TURING 图灵数学·统计学丛书 42

PEARSON





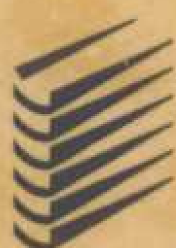
A First Course in Probability

概率论基础教程

（第8版）

[美] Sheldon M. Ross 著

郑忠国 詹从赞 译



人民邮电出版社
POSTS & TELECOM PRESS

概率论基础教程

A First Course in Probability

“这是一本优秀的概率论基础教材，是我所见到最好的一本。”

——Nhu Nguyen（新墨西哥州立大学）

“例子是如此地丰富和实用，写作风格清新、流畅，解答详细、准确，是一本很好读的教材……”

——Robert Bauer（伊利诺伊大学厄巴纳–尚佩恩分校）

概率论作为数学的一个重要分支，在众多领域发挥着越来越突出的作用。本书是全球高校采用率最高的概率论教材之一，初版于1976年，多年来不断重印修订，是作者几十年教学和研究经验的结晶。

本书叙述清晰，例子丰富，特别针对学生的兴趣选取了内容，有助于学生建立概率直觉。第8版与时俱进，增加了很多新的习题和例子，并新增两节内容，分别推导具有均匀分布和几何分布的随机变量和的分布。本书还附有大量习题、理论习题和自检习题，其中自检习题部分还给出全部解答，有利于读者巩固和自测所学知识。

Sheldon M. Ross 国际知名概率与统计学家，南加州大学工业工程与运筹系系主任。1968年博士毕业于斯坦福大学统计系，曾在加州大学伯克利分校任教多年。研究领域包括：随机模型、仿真模拟、统计分析、金融数学等。Ross教授著述颇丰，他的多种畅销数学和统计教材均产生了世界性的影响，其中*Simulation*（《统计模拟》）、*Introduction to Probability Models*（《应用随机过程：概率模型导论》）等均由人民邮电出版社引进出版。

延伸阅读

Bertsekas, Tsitsiklis 《概率导论（第2版）》

Feller 《概率论及其应用》，卷1和卷2

Jaynes 《概率论沉思录（英文版）》

DeGroot, Mark J. Schervish 《概率统计》

Ross 《应用随机过程：概率模型导论（第9版）》，中文版和英文版

Karlin, Taylor 《随机过程初级教程（第2版）》，中文版和英文版

Karlin, Taylor 《随机过程高级教程（英文版）》

Ash, Doleans-Dade 《概率与测度论（英文版·第2版）》

PEARSON

www.pearsonhighered.com





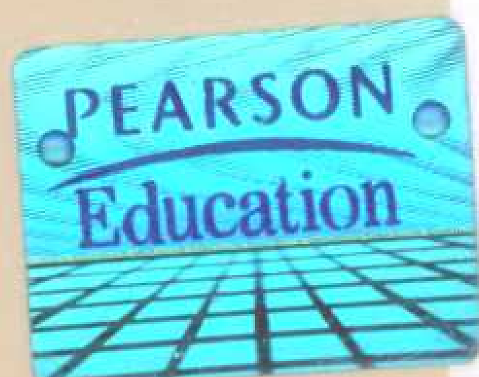
本书相关信息请访问：图灵网站 http://www.turingbook.com

读者/作者热线：(010)51095186

反馈/投稿/推荐信箱：contact@turingbook.com

分类建议：数学/概率

人民邮电出版社网址 www.ptpress.com.cn



ISBN 978-7-115-22110-0

定价：59.00 元

TURING 图灵数学·统计学丛书42

PEARSON



A First Course in Probability

概率论基础教程

（第8版）

[美] Sheldon M. Ross 著

郑忠国 詹从赞 译

人民邮电出版社

北京

图书在版编目(CIP)数据

概率论基础教程：第 8 版 /（美）罗斯（Ross, S. M.）著；郑忠国，詹从赞译．—北京：人民邮电出版社，2010.4
（图灵数学·统计学丛书）
书名原文：A First Course in Probability, Eighth Edition
ISBN 978-7-115-22110-0

Ⅰ．①概… Ⅱ．①罗…②郑…③詹… Ⅲ．①概率论教材 Ⅳ．①O211

中国版本图书馆 CIP 数据核字（2010）第 003805 号

内 容 提 要

概率论是研究自然界和人类社会中随机现象数量规律的数学分支．本书通过大量的例子讲述了概率论的基础知识，主要内容有组合分析、概率论公理化、条件概率和独立性、离散和连续型随机变量、随机变量的联合分布、期望的性质、极限定理等．本书附有大量的练习，分为习题、理论习题和自检习题三大类，其中自检习题部分还给出全部解答．

本书作为概率论的入门书，适用于大专院校数学、统计、工程和相关专业（包括计算科学、生物、社会科学和管理科学）的学生阅读，也可供应用工作者参考．

图灵数学·统计学丛书

概率论基础教程（第 8 版）

◆ 著　[美] Sheldon M. Ross
译　郑忠国　詹从赞
责任编辑　明永玲

◆ 人民邮电出版社出版发行　北京市崇文区夕照寺街 14 号
邮编 100061　电子函件 315@ptpress.com.cn
网址 http://www.ptpress.com.cn
北京鑫正大印刷有限公司印刷

◆ 开本：700×1000 1/16
印张：26.75
字数：548 千字　2010 年 4 月第 1 版
印数：1 -3 000 册　2010 年 4 月北京第 1 次印刷

著作权合同登记号　图字：01-2009-7274 号

ISBN 978-7-115-22110-0

定价：59.00 元

读者服务热线：(010)51095186　印装质量热线：(010) 67129223

反盗版热线：(010)67171154

版权声明

译 者 介 绍

郑忠国　北京大学数学科学学院教授、博士生导师, 1965 年北京大学研究生毕业. 长期从事数理统计的教学和科研工作, 研究方向是非参数统计、可靠性统计和统计计算, 发表论文近百篇. 主持完成国家科研项目“不完全数据统计理论及其应用”, 教育部博士点基金项目“应用统计方法研究”和“工业与医学中的应用统计研究”等. 研究项目“随机加权法”获国家教委科技进步二等奖. 出版的教材有《高等统计学》、《概率与统计》(北京大学出版社) 等.

詹从赞　毕业于北京大学概率统计专业. 毕业后一直从事于证券研究、产品设计等工作, 先后在指南针证券研究公司、金融界网站工作. 在《证券日报》等媒体发表文章若干, 曾任央视《今日证券》嘉宾. 著有《证券分析核心技术指标大全》、《不败而胜》等著作.

译者序

概率论是研究自然界和人类社会中随机现象数量规律的数学分支. 概率论的理论和方法与数学的其他分支、自然科学、工程、人文及社会科学各领域相互交叉渗透, 已经成为这些学科中的基本方法. 概率论 (或概率统计) 和高等数学一样已经成为我国高等院校各专业普遍设立的一门基础课.

目前, 这方面的教材已经很多, 但这本由 Sheldon M. Ross 编写的《概率论基础教程》确实是一本很有特点的好教材. 如在介绍概率的概念时, 作者还用流畅的笔调介绍了这些概念的发展历史, 从独立重复试验事件发生频率的极限到近代概率论的公理, 同时引用大量例子介绍如何利用概率的公理进行概率的计算. 这种讲法, 使得即使是只具有初等微积分知识的读者, 也会获益匪浅, 对概率的概念有一个正确和深刻的认识. 在介绍数学期望的概念时, 作者用大量的例子, 强调应用期望的性质, 特别是利用可加性进行期望计算, 从而使读者加深了对期望的认识, 也提高了运算技巧. 从本书第 1 章到第 8 章, 讲授的主题着重于概率论最基本的概念, 如概率、条件概率、期望、大数定律和中心极限定理等. 本书附有大量的有意义的练习, 分为习题、理论习题和自检习题三大类, 其中自检习题部分还给出全部解答, 以供参考. 从以上分析看出, 本书完全实现了作者在前言中所提的目标 —— 试图成为概率论的入门书.

本书第 1 版出版于 1976 年, 1981 年在国内曾出过第 1 版的中文翻译版. 此书经过作者历次修改, 内容大大扩充. 我们曾于 2006 年将原文第 7 版翻译成中文, 由人民邮电出版社出版. 此次, 作者又在第 7 版的基础上加以修订, 写成第 8 版. 第 8 版语言更加精炼, 并仔细斟酌了其中的例子. 因此, 本书是经过锤炼的优秀教材. 此外作者的另一本著作《随机过程》① 已经成为国内概率统计界推崇的教材. 我们相信本教材也一定会受到国内各界的欢迎.

由于译者的学识和中英文水平有限, 译文难免会有不妥之处, 欢迎广大读者批评指正.

译者谨识

2009 年 11 月

① 中国统计出版社, 2005. —— 编者注

前　言

法国著名数学家和天文学家拉普拉斯侯爵 (人称“法国的牛顿”) 曾经说过:“我们发现概率论其实就是将常识问题归结为计算. 它使我们能够精确地评价凭某种直观感受到的、往往又不能解释清楚的见解……值得注意的是, 概率论这门起源于机会游戏的科学, 早就应该成为人类知识中最重要的组成部分……生活中那些最重要的问题绝大部分恰恰是概率论问题. ”尽管许多人认为, 这位对概率论的发展作出过重大贡献的著名侯爵说话有点过头, 然而今日, 概率论已经成为几乎所有的科学工作者、工程师、医务人员、法律工作者以及企业家们手中的基本工具, 这是一个不争的事实. 事实上, 现代人们不再问:“是这样吗? ”而是问:“这件事发生的概率有多大? ”

本书试图成为概率论的入门书. 读者对象是数学、统计、工程和其他专业 (包括计算机科学、生物学、社会科学和管理科学) 的学生. 他们的先修知识只是初等微积分. 本书试图介绍概率论的数学理论, 同时通过大量例子说明这门学科的广泛的应用.

第 1 章介绍了组合分析的基本原理, 它是计算概率的最有效的工具.

第 2 章介绍了概率论的公理体系, 并且指出如何应用这些公理进行概率计算.

第 3 章讨论概率论中极为重要的概念, 即事件的条件概率和事件间的独立性. 通过一系列例子说明, 当部分信息可利用时, 条件概率就会发挥它的作用; 即使在没有这部分信息时, 条件概率也可以使概率的计算变得容易、可行. 利用“条件”计算概率这一极为重要的技巧还将出现在第 7 章, 在那里我们用它来计算期望.

在第 4~6 章, 我们引进随机变量的概念. 第 4 章讨论离散随机变量, 第 5 章讨论连续随机变量, 而将随机变量的联合分布放在第 6 章. 在第 4 章和第 5 章中讨论了随机变量的期望和方差, 并且对许多常见的随机变量, 求出了相应的期望和方差.

第 7 章讨论了期望值和它的一些重要的性质. 书中引入了许多例子, 解释如何利用随机变量和的期望等于随机变量期望的和这一重要规律来计算随机变量的期望, 本章中还有几节介绍条件期望 (包括它在预测方面的应用) 和矩母函数等. 最后一节介绍了多元正态分布, 同时给出了来自正态总体的样本均值和样本方差的联合分布的简单证明.

第 8 章介绍了概率论主要的理论结果. 特别地, 我们证明了强大数定律和中心极限定理. 关于强大数定律的证明, 我们假定随机变量具有有限的四阶矩. 在这种假定之下, 证明十分简单. 在中心极限定理的证明中, 我们假定了莱维连续性定理成立. 在本章中, 我们还介绍了若干概率不等式, 如马尔可夫不等式、切比雪夫不等式和切尔诺夫界. 在最后一节, 我们给出用随机变量的相应概率去近似独立伯努

利随机变量和的相关概率的误差界.

第 9 章介绍了一些附加主题, 如马尔可夫链、泊松过程以及信息编码理论初步. 第 10 章介绍了统计模拟.

与前几版一样, 每章后面附了三组练习题, 分别命名为**习题**、**理论习题**和**自检习题**. 在附录 B 中提供了自检习题的全部解答[①], 以供学生检验他们的理解能力.

新版中的新增或改变部分

第 8 版继续对原有教材作了微调和改进. 从培养学生的直观能力和学习兴趣出发, 新版增添了一些例子和习题. 例如, 在第 1 章例 5d 讨论了淘汰赛问题, 在第 7 章例 4k 和例 5i 讨论多人博奕的破产问题.

这一版的一个重要的改变是将 "随机变量和的期望等于各随机变量期望的和" 这一重要结论由原来 (第 7 版) 的第 7 章挪到第 4 章. 新版中为这个结论提供了新的初等证明 (有限样本空间的情况).

另一个改变是扩充了原 6.3 节的内容. 6.3.1 节是全新的内容. 在这一节中, 我们导出独立同分布的均匀随机变量的和的分布. 并且利用这个结果导出了这样的结果: 随机变量部分和超过 1 所需的项数的期望值为 e. 6.3.5 节也是新的. 这一节讨论一组独立几何随机变量的和的分布问题, 此处各随机变量参数 p_i 可以是不同的值. 这一节中导出了这个和的分布.

感谢

我感谢 Hossein Hamedani 对本书内容的仔细校对. 下列同仁为了改进教材, 提出了宝贵意见和建议, 在此一并表示感谢: Amir Ardestani (德黑兰理工大学), Joe Blitzstein (哈佛大学), Peter Nuesch (洛桑大学), Joseph Mitchell (纽约州立大学石溪分校), Alan Chambless (精算师), Robert Kriner, Israel David (本–古里安大学), T. Lim (乔治 · 梅森大学), Wei Chen (罗格斯大学), D. Monrad (伊利诺伊大学), W. Rosenberger (乔治 · 梅森大学), E. Ionides (密歇根大学), J. Corvino (拉法叶学院), T. Seppalainen (威斯康星大学).

我们还要感谢下列校阅者, 他们提出了宝贵的意见. 其中打 * 者为第 8 版的校阅者.

K. B. Athreya (爱荷华州立大学)
Richard Bass (康涅狄格大学)
Pobert Bauer (伊利诺伊大学厄巴纳–尚佩恩分校)
Phillip Beckwith (密歇根科技大学)
Arthur Benjamin (哈维姆德学院)
Geoffrey Berresford (长岛大学)
Baidurya Bhattacharya (特拉华大学)
Howard Bird (圣克劳德州立大学)

① 习题解答可从图灵网站下载. —— 编者注

Shahar Boneh (丹佛城市州立学院)
Jean Cadet (纽约州立大学石溪分校)
Steven Chiappari (圣克拉拉大学)
Nicolas Christou (加州大学洛杉矶分校)
James Clay (亚利桑那大学图森分校)
Francis Conlan (圣克拉拉大学)
*Justin Corvino (拉法叶学院)
Jay DeVore (圣路易斯–奥比斯波的加州技术大学)
Scott Emerson (华盛顿大学)
Thomas R.Fischer (德州农机大学)
Anant Godbole (密歇根科技大学)
Zakkhula Govindarajulu (肯塔基大学)
Richard Groeneveld (爱荷华州立大学)
Mike Hardy (麻省理工学院)
Bernard Harris (威斯康星大学)
Larry Harris (肯塔基大学)
David Heath (康奈尔大学)
Stephen Herschkorn (罗格斯大学)
Julia L.Higle (亚利桑那大学)
Mark Huber (杜克大学)
*Edward Ionides (密歇根大学)
Anastasia Ivanova (北卡罗来纳大学)
Hamid Jafarkhani (加州大学厄文分校)
Chuanshu Ji (北卡罗来纳大学 Chapel Hill 分校)
Robert Keener (密歇根大学)
Fred Leysieffer (佛罗里达州立大学)
Thomas Liggett (加州大学洛杉矶分校)
Helmut Mayer (佐治亚大学)
Bill McCormick (佐治亚大学)
Ian McKeague (佛罗里达州立大学)
R. Miller (斯坦福大学)
*Ditlev Monrad (伊利诺伊大学)
Robb J. Muirhead (密歇根大学)
Joe Naus (罗格斯大学)
Nhu Nguyen (新墨西哥州立大学)
Ellen O'Brien (乔治 · 梅森大学)
N. U. Prabhu (康奈尔大学)

Kathryn Prewitt (亚利桑那州立大学)
Jim Propp (威斯康星大学)
*William F. Rosenberger (乔治 · 梅森大学)
Myra Samuels (普度大学)
I.R Savage (耶鲁大学)
Art Schwartz (密歇根大学安阿伯分校)
Therese Shelton (西南大学)
Malcolm Sherman (纽约州立大学奥尔巴尼分校)
Murad Taqqu (波士顿大学)
Eli Upfal (布朗大学)
Ed Wheeler (田纳西大学)
Allen Webster (布拉德利大学)

S. R.
smross@usc.edu

目　录

第 1 章　组合分析

1.1　引　言

首先, 我们提出一个与概率论有关的有趣的经典问题: 一个通信系统含 n 个天线, 顺序地排成一排, 只要没有两个连续的天线都失效, 那么这个系统就可以接收到信号, 此时称这个通信系统是*有效的*. 已经探明这 n 个天线里, 恰好有 m 个天线是失效的, 问此通信系统仍然有效的概率是多大? 举例来说, 设 $n = 4, m = 2$, 通信系统是否有效取决于这 n 个天线的设置方式 (它们的排列次序). 这 4 个天线一共有 6 种可能的设置方式

$$0\,1\,1\,0 \quad 0\,1\,0\,1 \quad 1\,0\,1\,0 \quad 0\,0\,1\,1 \quad 1\,0\,0\,1 \quad 1\,1\,0\,0$$

其中, 1 表示天线有效, 0 表示天线失效. 可以看出前 3 种情况下整个通信系统仍然有效, 而后 3 种情况下系统将失效, 因此, 若天线的设置方式是随机排列的, 所求的概率应该是 $\frac{3}{6} = \frac{1}{2}$. 对于一般的 n 和 m 来说, 用类似上述方法可以计算出所求概率. 即先计算使得系统仍有效的设置方式有多少种, 再计算总共有多少种设置方式, 两者相除即为所求概率.

从上所述可看出, 一个有效地计算事件发生结果数目的方法是非常有用的. 事实上, 概率论里的很多问题只要通过计算一个事件发生结果的数目就能得以解决. 关于计数的数学理论通常称为**组合分析**(combinatorial analysis).

1.2　计数基本法则

对我们的整个讨论来说, 以下关于计数的法则是基本的. 粗浅地说, 若一个试验有 m 个可能结果, 另一个试验有 n 个可能结果, 则两个试验一共有 mn 个结果.

计数基本法则

有两个试验, 其中试验 1 有 m 种可能发生的结果, 对应于试验 1 的每一个结果, 试验 2 有 n 种可能发生的结果, 则对这两个试验来说, 一共有 mn 种可能结果.

基本法则的证明　通过列举两个试验所有可能的结果来证明这个问题, 结果

如下:

$$\begin{matrix} (1,1) & (1,2) & \cdots & (1,n) \\ (2,1) & (2,2) & \cdots & (2,n) \\ \vdots & \vdots & \vdots & \vdots \\ (m,1) & (m,2) & \cdots & (m,n) \end{matrix}$$

其中, (i,j) 表示第一个试验结果是第 i 种、第二个试验结果是第 j 种. 因此, 所有可能结果组成一个矩阵, 共有 m 行 n 列, 结果的总数为 $m \times n$, 这样就完成了证明.

例 2a 一个小团体由 10 位妇女组成, 每位妇女又有 3 个孩子. 现在要从其中选取一位妇女和她的一个孩子评为"年度母亲和年度儿童", 问一共有多少种可能的选取方式?

解: 将选择妇女看成第一个试验, 而接下来选择这位母亲的一个孩子看作第二个试验, 那么根据计数基本法则可知, 一共有 $10 \times 3 = 30$ 种选择方式. ■

当有 2 个以上的试验时, 基本法则可以推广如下.

推广的计数基本法则

一共有 r 个试验. 第一个试验有 n_1 种可能结果; 对应于第一个试验的每一种试验结果, 第二个试验有 n_2 种可能结果; 对应于头两个试验的每一种试验结果, 第三个试验有 n_3 种可能结果; 等等. 那么, 这 r 个试验一共有 $n_1 \cdot n_2 \cdots n_r$ 种可能结果.

例 2b 一个大学计划委员会由 3 名新生、4 名二年级学生、5 名三年级学生、2 名毕业班学生组成. 现在要从中选 4 个人组成一个分委员会, 并要求分委员会的成员来自不同的年级, 一共有多少种选择方式?

解: 可以把它理解为从每个年级选取一个代表, 从而有 4 个试验, 根据推广的计数基本法则, 一共有 $3 \times 4 \times 5 \times 2 = 120$ 种可能的选择结果. ■

例 2c 车牌号是 7 位的, 如果要求前 3 个位置必须是字母, 后 4 个必须是数字, 一共有多少种编排车牌号的方式?

解: 根据推广的计数基本法则, 可知道答案为: $26 \times 26 \times 26 \times 10 \times 10 \times 10 \times 10 = 175\,760\,000$. ■

例 2d 对于只定义在 n 个点上的函数, 如果函数取值只能为 0 或 1, 这样的函数有多少?

解: 设这 n 个点为 $1, 2, \cdots, n$, 既然对每个点来说, $f(i)$ 的取值只能为 0 或者 1, 那么一共有 2^n 个这样的函数. ■

例 2e 在例 2c 中, 如果不允许字母或数字重复, 一共有多少种可能的车牌号?

解: 这种情况下, 一共有 $26 \times 25 \times 24 \times 10 \times 9 \times 8 \times 7 = 78\,624\,000$ 种可能的车牌号. ■

1.3 排 列

按随意顺序来排列字母 a, b, c, 一共有多少种排列方式? 通过直接列举, 可知一共有 6 种: abc, acb, bac, bca, cab 以及 cba. 每一种都可以称为一个*排列* (permutation). 因此, 3 个元素一共有 6 种可能排列方式. 这个结果能通过推广的计数基本法则得到: 在排列中第一个位置可供选择的元素有 3 个, 第二个位置可供选择的元素是剩下的两个之一, 第三个位置只能选择剩下的 1 个元素, 因此一共有 $3\times2\times1=6$ 种可能的排列.

假设有 n 个元素, 那么用上述类似的方法, 可知一共有 $n(n-1)(n-2)\cdots3\cdot2\cdot1=n!$ 种不同的排列方式.

例 3a 一个垒球队一共有 9 名队员, 问一共有多少种击球顺序?

解: 一共有 $9!=362\,880$ 种可能的击球顺序. ■

例 3b 某概率论班共有 6 名男生、4 名女生, 对班上的学生进行一次测验, 并根据测验成绩排名次, 假设没有两个学生成绩一样.

(a) 一共有多少种排名次的方式?

(b) 如限定男生、女生分开排名次, 一共有多少种排名次的方式?

解:

(a) 每种排名方法都对应着一个 10 人的排列方式, 故答案是: $10!=3\,628\,800$.

(b) 男生一起排名次有 $6!$ 种可能, 女生一起排名次有 $4!$ 种, 根据计数基本法则, 一共有 $6!\times4!=720\times24=17\,280$ 种可能结果. ■

例 3c 把 10 本书放到书架上, 其中有 4 本数学书、3 本化学书、2 本历史书和 1 本语文书. 现在要求相同类别的书必须紧挨着放, 问一共有多少种放法?

解: 如果数学书放在最前面, 接下来放化学书, 再下来放历史书, 最后放语文书, 那么一共有 $4!3!2!1!$ 种排列方式. 而这 4 种书的顺序一共又是 $4!$ 种, 因此, 所求答案是 $4!4!3!2!1!=6912$. ■

接下来讨论如果有 n 个元素, 其中有些是不可区分的, 这种排列数如何计算? 先看下面的例子.

例 3d 用 PEPPER 的 6 个字母进行排列, 一共有几种不同的排列方式?

解: 如果3个字母 P 和 2 个字母 E 都是可以区分的 (标上号), 即 $P_1E_1P_2P_3E_2R$, 一共有 $6!$ 种排列方式. 然而, 考察其中任一个排列, 比如 $P_1P_2E_1P_3E_2R$, 如果分别将 3 个字母 P 和 2 个字母 E 的次序重排, 那么得到的结果仍然是 PPEPER, 也就是说, 总共有 $3!2!$ 种排列

$$
\begin{array}{llll}
P_1P_2E_1P_3E_2R & P_1P_2E_2P_3E_1R & P_1P_3E_1P_2E_2R & P_1P_3E_2P_2E_1R \\
P_2P_1E_1P_3E_2R & P_2P_1E_2P_3E_1R & P_2P_3E_1P_1E_2R & P_2P_3E_2P_1E_1R \\
P_3P_1E_1P_2E_2R & P_3P_1E_2P_2E_1R & P_3P_2E_1P_1E_2R & P_3P_2E_2P_1E_1R
\end{array}
$$

这些排列都具有形式: PPEPER. 因此一共有 $6!/(3!2!)=60$ 种不同的排列方式. ■

一般来说, 利用上述同样的推论方法可知: n 个元素, 如果其中 n_1 个元素彼此不可区分, 另 n_2 个彼此不可区分, $\cdots$, n_r 个也彼此不可区分, 那么一共有 $\dfrac{n!}{n_1!n_2!\cdots n_r!}$ 种排列方式.

例 3e 一个棋类比赛一共有 10 个选手, 其中 4 个来自俄罗斯, 3 个来自美国, 2 个来自英国, 另 1 个来自巴西. 如果比赛结果只记录选手的国籍, 那么一共有多少种可能结果?

解: 一共有 $\dfrac{10!}{4!3!2!1!}=12\,600$ 种可能结果. ■

例 3f 有 9 面小旗排列在一条直线上, 其中 4 面白色、3 面红色和 2 面蓝色, 颜色相同的旗是一样的. 如果不同的排列方式代表不同的信号, 那么一共有多少种可能的信号?

解: 一共有 $\dfrac{9!}{4!3!2!}=1260$ 种不同的信号. ■

1.4 组 合

从 n 个元素当中取 r 个, 一共有多少种取法? 这也是一个有趣的问题. 比如, 从 A, B, C, D 和 E 这 5 个元素中选取 3 个组成一组, 一共有多少种取法? 解答如下: 取第一个有 5 种取法, 取第 2 个有 4 种取法, 取第三个有 3 种取法, 所以, 如果考虑选择顺序的话, 那么一共有 $5\times4\times3$ 种取法. 但是, 每一个包含 3 个元素的组 (比如包含 A, B, C 的组) 都被计算了 6 次, (也即, 如果考虑顺序的话, 所有的排列 ABC, ACB, BAC, BCA, CAB, CBA 都被算了一次.) 所以, 组成方法数为:

$$\frac{5\times4\times3}{3\times2\times1}=10$$

一般来说, 如果考虑顺序的话, 从 n 个元素中选择 r 个组成一组一共有 $n(n-1)\cdots(n-r+1)$ 种方式, 而每个含 r 个元素的小组都被重复计算了共 $r!$ 次. 所以, 能组成不同的组的数目为:

$$\frac{n(n-1)\cdots(n-r+1)}{r!}=\frac{n!}{(n-r)!r!}$$

记号与术语

对 $r\leqslant n$, 定义 $\binom{n}{r}$ 如下:

$$\binom{n}{r}=\frac{n!}{(n-r)!\,r!}$$

并且说 $\binom{n}{r}$ 表示了从 n 个元素中一次取 r 个的可能组合数.[1]

① 为了方便, 0! 被定义为 1, 因此, $\binom{n}{0}=\binom{n}{n}=1$. 当 $i<0$ 或者 $i>n$ 时, 有时也认为 $\binom{n}{i}$ 等于 0.

因此, $\binom{n}{r}$ 就表示了从 n 个元素中一次取 r 个元素的可能取法的数目, 如果不考虑抽取顺序的话.

例 4a 从 20 人当中选择 3 人组成委员会, 一共有多少种选法?

解: 一共有 $\binom{20}{3}=\dfrac{20\times 19\times 18}{3\times 2\times 1}=1140$ 种选法. ■

例 4b 有个 12 人组成的团体, 其中 5 位女士, 7 位男士, 现从中选取 2 位女士, 3 位男士组成一个委员会, 问有多少种取法? 另外, 如果其中有 2 位男士之间有矛盾, 并且坚决拒绝一起工作, 那又有多少种取法?

解: 有 $\binom{5}{2}$ 种方法选取女士, 有 $\binom{7}{3}$ 种方法选取男士, 根据基本计数法则一共有 $\binom{5}{2}\binom{7}{3}=\dfrac{5\times 4}{2\times 1}\times\dfrac{7\times 6\times 5}{3\times 2\times 1}=350$ 种方式选取 2 位女士 3 位男士.

现在来看如果有两位男士拒绝一起工作, 那么选取 3 位男士的 $\binom{7}{3}=35$ 种方法中, 有 $\binom{2}{2}\binom{5}{1}=5$ 种同时包含了该两位男士, 所以, 一共有 $35-5=30$ 种选取方法不同时包含那两位有矛盾的男士; 另外, 选取女士的方法仍是 $\binom{5}{2}=10$ 种, 所以, 一共有 $30\times 10=300$ 种选取方式. ■

例 4c 假设一排 n 个天线中, 有 m 个是失效的, 另 $n-m$ 个是有效的, 并且假设所有有效的天线之间不可区分, 同样, 所有失效的天线之间也不可区分. 问有多少种排列方式, 使得没有两个连续的天线是失效的?

解: 先将 $n-m$ 个有效天线排成一排, 既然没有连续两个失效的, 那么两个有效天线之间, 必然至多放置一个失效的. 也即, 在 $n-m+1$ 个可能位置中 (如图 1.1 中的星号), 选择 m 个来放置失效天线. 因此有 $\binom{n-m+1}{m}$ 种可能方式确保在两个失效天线之间至少存在一个有效天线.[①] ■

以下是一个非常有用的组合恒等式:

$$\binom{n}{r}=\binom{n-1}{r-1}+\binom{n-1}{r}\qquad 1\leqslant r\leqslant n \tag{4.1}$$

上述恒等式可用分析的方法证明, 也可从组合的角度来证明. 设想从 n 个元素中取 r 个, 一共有 $\binom{n}{r}$ 种取法. 从另一个角度来考虑, 不妨设这 n 个元素里有一个特殊的, 记为元素 1, 那么取 r 个元素就有两种结果, 取元素 1 或者不取元素 1. 取元素

① 若 $m>n-m+1$, 按组合记号的约定, $\binom{n-m+1}{m}=0$. 这也符合实际, 当失效天线太多的时候, 任何天线的次序排列, 总有两个相邻的天线是失效的. —— 译者注

1 的方法一共有 $\binom{n-1}{r-1}$ 种 (从 $n-1$ 个元素里面取 $r-1$ 个); 不取元素 1 的方法一共有 $\binom{n-1}{r}$ 种 (从去掉元素 1 的剩下 $n-1$ 个元素中取 r 个). 两者之和就是从 n 个元素里取 r 个的方法之和, 所以恒等式成立.

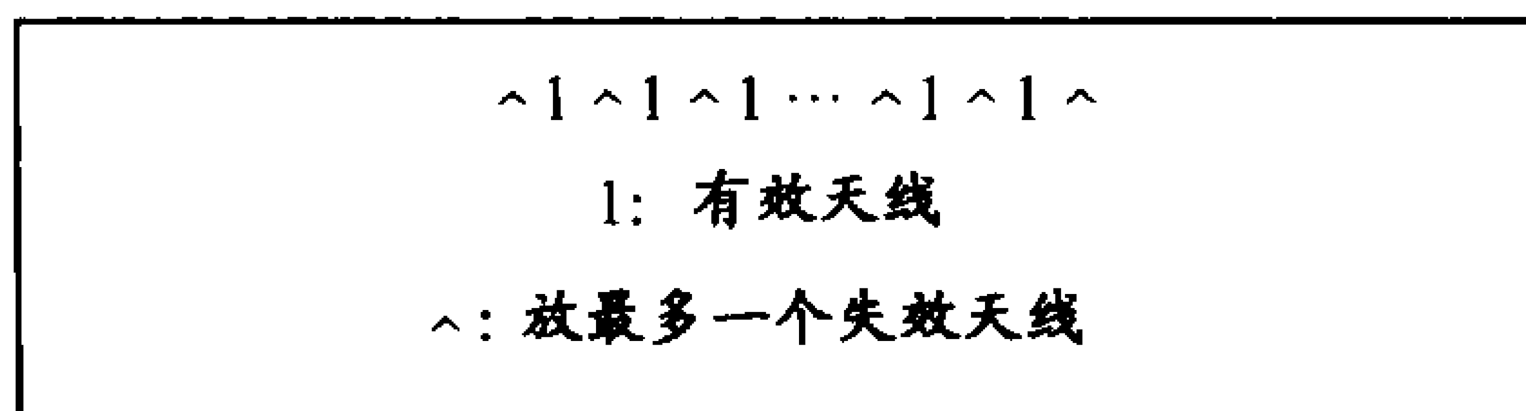


图 1.1　天线的排列

值 $\binom{n}{r}$ 经常也称为二项式系数(binomial coefficient), 这是因为它们是以下的二项式定理中重要的系数.

二项式定理

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} \tag{4.2}$$

以下提供二项式定理的两个证明方法, 其一是数学归纳法, 其二是基于组合考虑的证明.

二项式定理的归纳法证明　$n=1$ 时, (4.2) 式化为 $x+y = \binom{1}{0}x^0y^1 + \binom{1}{1}x^1y^0 = y+x$. 现假设 (4.2) 式对于 $n-1$ 成立, 那么对于 n,

$$\begin{aligned}(x+y)^n &= (x+y)(x+y)^{n-1} = (x+y)\sum_{k=0}^{n-1}\binom{n-1}{k}x^k y^{n-1-k} \\ &= \sum_{k=0}^{n-1}\binom{n-1}{k}x^{k+1}y^{n-1-k} + \sum_{k=0}^{n-1}\binom{n-1}{k}x^k y^{n-k}\end{aligned}$$

在前面的求和公式里令 $i=k+1$, 后面的求和公式里令 $i=k$, 那么

$$\begin{aligned}(x+y)^n &= \sum_{i=1}^{n}\binom{n-1}{i-1}x^i y^{n-i} + \sum_{i=0}^{n-1}\binom{n-1}{i}x^i y^{n-i} \\ &= x^n + \sum_{i=1}^{n-1}\left[\binom{n-1}{i-1} + \binom{n-1}{i}\right]x^i y^{n-i} + y^n \\ &= x^n + \sum_{i=1}^{n-1}\binom{n}{i}x^i y^{n-i} + y^n = \sum_{i=0}^{n}\binom{n}{i}x^i y^{n-i}\end{aligned}$$

这样就证明了等式. ■

二项式定理的组合法证明　考虑乘积 $(x_1+y_1)(x_2+y_2)\cdots(x_n+y_n)$ 展开后一共包含 2^n 项, 每一项都是 n 个因子的乘积, 而且每一项都包含因子 x_i 或 y_i, 例如:

$$(x_1+y_1)(x_2+y_2)=x_1x_2+x_1y_2+y_1x_2+y_1y_2$$

这 2^n 项和里面, 一共有多少项含有 k 个 x_i 和 $n-k$ 个 y_i?

含有 k 个 x_i 和 $n-k$ 个 y_i 的每一项对应了从 n 个元素 $x_1,x_2,\cdots,x_n$ 里取 k 个元素的取法. 因此一共有 $\binom{n}{k}$ 个这样的项. 这样, 令 $x_i=x, y_i=y, i=1,\cdots,n$, 可以看出

$$(x+y)^n=\sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k}$$ ■

例 4d 展开 $(x+y)^3$.

解:
$$\begin{aligned}(x+y)^3&=\binom{3}{0}x^0y^3+\binom{3}{1}x^1y^2+\binom{3}{2}x^2y+\binom{3}{3}x^3y^0\\&=y^3+3xy^2+3x^2y+x^3\end{aligned}$$ ■

例 4e 一个有 n 个元素的集合共有多少子集?

解: 含有 k 个元素的子集一共有 $\binom{n}{k}$ 个, 因此所求答案为

$$\sum_{k=0}^{n}\binom{n}{k}=(1+1)^n=2^n.$$

该结果还可以这样得到: 给该集合里的每个元素都标上 1 或 0, 每种标法都一一对应了一个子集, 例如, 当把所有元素都标为 1 时候, 就对应着一个含有所有元素的子集. 因为一共有 2^n 种标法, 所以一共有 2^n 个子集.

上述结论包含了一个元素都没有的子集 (也即空集), 所以至少有一个元素的子集一共有 2^n-1 个. ■

1.5 多项式系数

本节考虑如下问题: 有 n 个不同的元素, 分成 r 组, 每组分别有 $n_1,n_2,\cdots,n_r$ 个元素, 其中 $\sum_{i=1}^{r}n_i=n$, 一共有多少种分法? 注意到, 第一组成员有 $\binom{n}{n_1}$ 种选取方法, 接下来, 选定第一组成员后, 选第二组成员时只能从剩下的 $n-n_1$ 个元素中选, 一共有 $\binom{n-n_1}{n_2}$ 种取法, 接下来第三组有 $\binom{n-n_1-n_2}{n_3}$ 种取法, 等等. 因此, 根据推广的计数基本法则, 将 n 个元素分成 r 组的分法总数一共是

$$\begin{aligned}&\binom{n}{n_1}\binom{n-n_1}{n_2}\cdots\binom{n-n_1-n_2-\cdots-n_{r-1}}{n_r}\\&=\frac{n!}{(n-n_1)!\,n_1!}\cdot\frac{(n-n_1)!}{(n-n_1-n_2)!\,n_2!}\cdots\frac{(n-n_1-n_2-\cdots-n_{r-1})!}{0!\,n_r!}\\&=\frac{n!}{n_1!\,n_2!\cdots n_r!}\end{aligned}$$

用另一个方法也可以得到此结果：有 n 个球，其中 n_1 个标有记号 1，n_2 个标有记号 2，$\cdots$，n_r 个标有记号 r. 具有相同标号的球之间是不可分辨的. 现在设有 n 个对象，记为 $1,2,\cdots,n$. 设想一个球里也只可放一个对象. 把 n 个对象放到 r 个球里边，等价于将 n 个对象分成 r 个组，因为每个球有一个标号，这个标号就是这个对象的组号. 现在问一共有多少不同的分组方式呢？先把这些对象排成一个固定的顺序，然后将这些球进行排列. 当球的排列确定以后，每一个对象就对应一个球的编号. 这就等价于将 n 个对象分成 r 个组，每个对象的组号就是这个对象所对应的球号. 这些排列数就是这 n 个对象的分组数. 在 1.3 节里讨论了 n 个对象的排列数的计数方法，其中具有相同球号的球之间是不可分辨的. 在这种情况下，n 个球的排列数为 $\dfrac{n!}{n_1!\cdots n_r!}$.

记号

如果 $n_1+n_2+\cdots+n_r=n$，定义 $\binom{n}{n_1,n_2,\cdots,n_r}$ 为

$$\binom{n}{n_1,n_2,\cdots,n_r}=\frac{n!}{n_1!n_2!\cdots n_r!}$$

因此，$\binom{n}{n_1,n_2,\cdots,n_r}$ 表示了把 n 个不同的元素分成大小分别为 $n_1,n_2,\cdots,n_r$ 的 r 组的方法数.[①]

例 5a 某小城市的警察局有 10 名警察，其中需要 5 名警察在街道巡逻，2 名警察在局里值班，另 3 名留在局里待命. 问把 10 名警察分成这样的 3 组共有多少种不同分法？

解：一共有 $10!/(5!2!3!)=2520$ 种分法. ■

例 5b 10 个小孩要分成 A,B 两队，每队 5 人. A 队去参加一场比赛，B 队则去参加另一场比赛，一共有多少种分法？

解：一共有 $10!/(5!5!)=252$ 种分法. ■

例 5c 10 个孩子分成两组，每组 5 人进行篮球比赛，一共有多少种分法？

解：这个问题与例 5b 不同的是，分成的两组是不用考虑顺序的. 也就是说，这儿没有 A,B 两组之分，仅仅分成各自为 5 人的两组，故所求答案为

$$\frac{10!/(5!5!)}{2!}=126$$ ■

下面的定理是二项式定理的推广，其证明留作习题.

① 为了表达方便，今后将 n 个不同元素分成大小不同的“组”称为“多项组合”，分组的方法数 $\binom{n}{n_1,\cdots,n_r}$ 称为“多项组合数”. —— 译者注

多项式定理

$$(x_1+x_2+\cdots+x_r)^n=\sum_{\substack{(n_1,\cdots,n_r):\\ n_1+\cdots+n_r=n}}\binom{n}{n_1,n_2,\cdots,n_r}x_1^{n_1}x_2^{n_2}\cdots x_r^{n_r}$$

上式的求和号是对一切满足 $n_1+n_2+\cdots+n_r=n$ 的所有非负整向量 $(n_1,n_2,\cdots,n_r)$ 求和.

鉴于多项式定理的形式, 多项组合数 $\binom{n}{n_1,n_2,\cdots,n_r}$ 也称为**多项式系数** (multinomial coefficient).

例 5d 考虑涉及 $n=2^m$ 个人的淘汰赛. 将这 n 个参赛者随机地分成 $\frac{n}{2}$ 对, 并且让每一对选手进行比赛, 比赛的赢者留下来, 进入下一轮的比赛, 而失败者则淘汰出局. 这样, 一轮一轮比赛下去, 直到最后一个得胜者, 成为冠军. (假设淘汰赛中一共有 8 个参赛者.)

(a) 在第一轮比赛中, 有多少不同的可能结果? (例如: 其中一个结果为 1 战胜 2, 3 战胜 4, 5 战胜 6, 7 战胜 8.)

(b) 在整个淘汰赛中, 一共有多少种不同的结果, 其中每一个结果展示了各轮比赛的全部信息.

解: 可以通过确定第一轮比赛的可能分组方法数来确定第一轮比赛的可能结果数. 注意到, 将 8 个参赛者分成第一对、第二对、第三对、第四对, 共有 $\binom{8}{2,2,2,2}=\frac{8!}{4!}$ 种分法. 因此, 当不考虑这 4 对的先后顺序时, 可能的分组方法数是 $\frac{8!}{2^4 4!}$. 对于每一种分组方法中的每一对选有两种比赛结果, 这表明第一轮比赛中一共有 $\frac{8!2^4}{2^4 4!}=\frac{8!}{4!}$ 种可能的结果.①

① 为了具体化, 我们将淘汰赛理解为由 8 个选手组成的乒乓球淘汰赛. 假定第一轮比赛中一共有 4 张桌子, 并引入几个概念. 将 8 名运动员分配到 4 张乒乓桌进行比赛, 每种分配方案称为一个 "分配", 例如 $[(1,2),(3,4),(5,6),(7,8)]$ 表示一个分配, 将 1 号和 2 号运动员分配在第 1 桌, $\cdots$, 将 7 号和 8 号运动员分配在第 4 桌. 分配方案数就是多项组合数 $\binom{8}{2,2,2,2}=\frac{8!}{2^4}$. 我们引入的第二个概念就是 "分组". 将 1 号和 2 号运动员分在一个组进行比赛, $\cdots$, 将 7 号和 8 号运动员分在一组进行比赛, 而不管这些运动员的比赛是在哪一张桌子进行的, 例如 $[(1,2),(3,4),(5,6),(7,8)]$ 与 $[(3,4),(1,2),(5,6),(7,8)]$ 是不同的分配, 但从分组的角度看, 两者没有区别, 属于同一个分组方案. 分配方案数与分组方案数之间具有如下关系:

$$\text{分配方案数} = \text{分组方案数} \times 4!$$

因此, 在第一轮比赛中一共有 $\frac{8!}{2^4 4!}$ 种不同的分组方法. 对于每一种分组方法中的每一对选手有两种比赛结果. 这样, 每一种分组方法一共有 2^4 种不同的比赛结果. 利用计数基本法则可知, 在第一轮比赛中一共有 $\frac{8!2^4}{2^4 4!}=\frac{8!}{4!}$ 种可能结果.

[也可用另一种看法来理解这个结果：在 8 个选手中有 4 个得胜者. 这种得胜者的组合数为 $\binom{8}{4}$. 每一个得胜者组合 (例如, (1,2,3,4) 为得胜者之组合) 可有 4! 种不同的方式与 4 个失败者进行比赛, 这样在第一轮比赛中, 一共有 $4!\binom{8}{4}=\frac{8!}{4!}$ 个可能的比赛结果.]

类似地, 对于第一轮比赛的每一个可能结果, 其第二轮比赛的可能结果数为 $\frac{4!}{2!}$. 对于前面两轮比赛的每一种可能结果, 在第三轮比赛的结果数为 $\frac{2!}{1!}$. 利用推广的计数基本法则可知, 整个淘汰赛的所有可能的结果数为 $\frac{8!}{4!}\frac{4!}{2!}\frac{2!}{1!}=8!$. 类似的推论可知, 在 $n=2^m$ 个参赛者组成的淘汰赛中, 一共有 $n!$ 种可能的比赛结果.

知道了上面的结论, 我们很容易将淘汰赛的各种可能结果与 $(1,2,\cdots,n)$ 的各种排列之间建立一一对应关系. 对每一个比赛的可能结果, 给参加比赛者一个排名. 比赛的冠军排名为 1, 在最后决赛时, 输给冠军的人排名为 2. 在倒数第二轮中的两位输者的排名如下：与冠军比赛者的排名为 3, 与亚军比赛者排名为 4. 这样, 在倒数第三轮比赛中 4 名胜利者已经排出 1, 2, 3, 4 的名次. 这轮比赛中 4 名失败者排名为 5, 6, 7, 8. 与第 1 名比赛者排名 5, 依次下去, 与第 4 名比赛者排名 8. 这样继续排下去, 直到所有参赛者都得到排名. (可以用更简洁的方式描述这种名次. 淘汰赛的最后得胜者排名为 1, 其余的参赛者的名次是这样确定的：参赛者被打败的那一轮中的比赛场次 2^k, 再加上打败他的那个人的名次, $k=0,1,2,\cdots,m-1$.) 这样, 淘汰赛的结果可用排列 $(i_1,\cdots,i_m)$ 来表示, 其中 i_j 代表第 j 名参赛者的编号. 由于不同的淘汰赛的结果给出不同的排列, 而对于每一个排列, 都有一个淘汰赛的结果与之对应. 这样, 可能的淘汰赛的结果数目与 $1,\cdots,n$ 的排列数相同, 即 $n!$. ■

例 5e

$$\begin{aligned}(x_1+x_2+x_3)^2 &= \binom{2}{2,0,0}x_1^2x_2^0x_3^0+\binom{2}{0,2,0}x_1^0x_2^2x_3^0+\binom{2}{0,0,2}x_1^0x_2^0x_3^2\\ &\quad+\binom{2}{1,1,0}x_1^1x_2^1x_3^0+\binom{2}{1,0,1}x_1^1x_2^0x_3^1+\binom{2}{0,1,1}x_1^0x_2^1x_3^1\\ &= x_1^2+x_2^2+x_3^2+2x_1x_2+2x_1x_3+2x_2x_3\end{aligned}$$ ■

*1.6 方程的整数解个数①

把 n 个可分辨的球分到 r 个可分辨的坛子里, 一共有 r^n 种方式. 这是因为任一个球都有可能放到 r 个坛子中的任一个. 现在假设这 n 个球是不可分辨的,

① 此处或以后打 * 号表示这些材料是可以选读的, 或是选作的题目.

这种情况下一共有多少种可能结果呢? 由于球是不可分辨的, 将 n 个球分到 r 个坛子里的结果可以描述为向量 $(x_1, x_2, \cdots, x_r)$, 其中 x_i 表示分到第 i 个坛子里球的数量. 因此, 此问题也等同于求出满足 $x_1 + x_2 + \cdots + x_r = n$ 的非负整数向量 $(x_1, x_2, \cdots, x_r)$ 的个数. 为了计算它, 先考虑该方程的正整数解个数: 设想有 n 个不可分辨的球排成一排, 现把它们分成 r 组, 每组都不空. 要做到这一点, 我们可在 n 个球之间的 $n-1$ 个空隙里选取 $r-1$ 个 (如图 1.2 所示). 其证明留作习题.

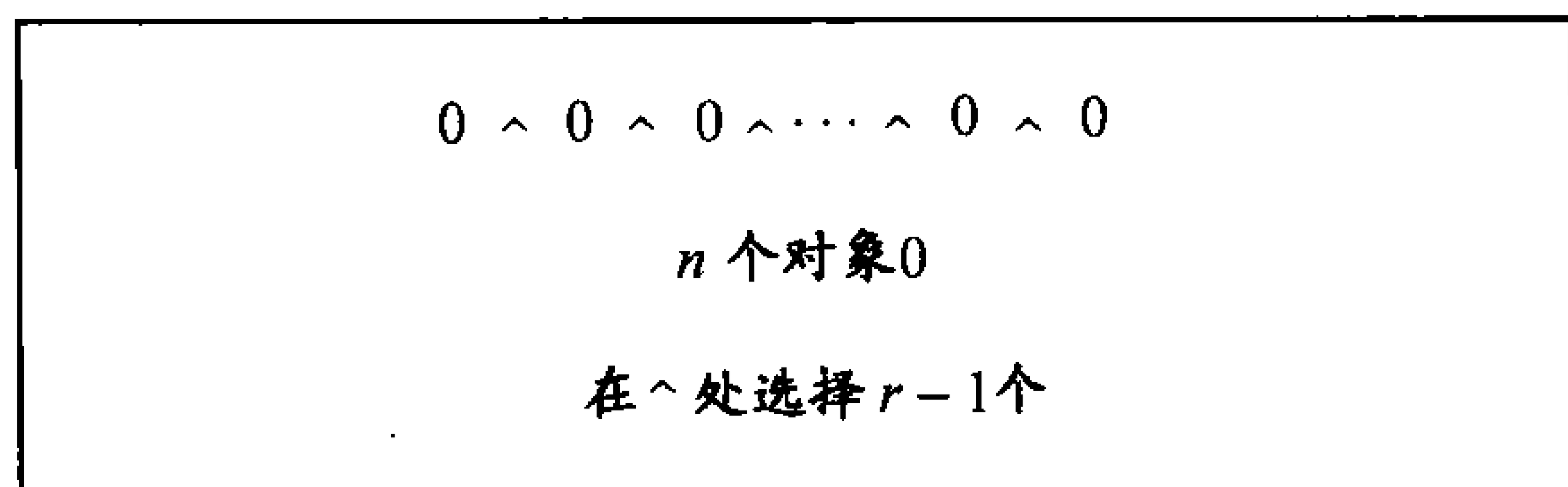


图 1.2 n 个球之间的 $n-1$ 个空隙里选取 $r-1$ 个

例如, $n = 8, r = 3$ 时, 可以选两个就能隔开:

$$000|000|00$$

代表向量 $x_1 = 3, x_2 = 3, x_3 = 2$. 由于一共有 $\binom{n-1}{r-1}$ 种选择可能, 因此就可以得到以下命题.

命题 6.1 共有 $\binom{n-1}{r-1}$ 个不同的具有正整数分量的向量 $(x_1, x_2, \cdots, x_r)$ 满足

$$x_1 + x_2 + \cdots + x_r = n \qquad x_i > 0, i = 1, \cdots, r$$

为了得到非负整数解 (而不是正整数解) 个数, 注意到 $x_1 + x_2 + \cdots + x_r = n$ 的非负整数解个数与 $y_1 + y_2 + \cdots + y_r = n + r$ 的正整数解个数是相同的 (令 $y_i = x_i + 1, i = 1, \cdots, r$). 因此, 利用命题 6.1, 可得到如下命题.

命题 6.2 共有 $\binom{n+r-1}{r-1}$ 个不同的具有非负整数分量的向量 $(x_1, x_2, \cdots, x_r)$ 满足

$$x_1 + x_2 + \cdots + x_r = n \tag{6.1}$$

例 6a 方程 $x_1 + x_2 = 3$ 共有多少组不同的非负整数解?

解: 一共有 $\binom{3+2-1}{2-1} = 4$ 组解: $(0,3), (1,2), (2,1), (3,0)$. ■

例 6b 有 2 万美元可投资到 4 个项目上, 每份投资必须是 1000 美元的整数倍. 如果要求 2 万美元全部投资, 一共有多少种可行的投资方法? 如果不要求将钱全部投资呢?

解: 令 x_i $(i=1,2,3,4)$ 分别表示 4 个项目的投资额 (单位: 千美元), 因此, 4 个项目的投资额就是方程 $x_1+x_2+x_3+x_4=20$ $x_i \geqslant 0$ 的非负整数解. 根据命题 6.2, 一共有 $\binom{23}{3}=1771$ 种可能的投资方式. 如果并不需要将钱全部投资, 那么假设 x_5 表示剩余资金, 那么一个投资策略就对应了如下方程的一个非负整数解:

$$x_1+x_2+x_3+x_4+x_5=20$$

因此, 根据命题 6.2, 存在 $\binom{24}{4}=10\ 626$ 种投资策略. ■

例 6c 在 $(x_1+x_2+\cdots+x_r)^n$ 的展开式中, 一共有多少项?

解:
$$(x_1+x_2+\cdots+x_r)^n=\sum\binom{n}{n_1,\cdots,n_r}x_1^{n_1}\cdots x_r^{n_r}$$

其中, 求和针对所有满足 $n_1+\cdots+n_r=n$ 的非负整数 $(n_1,\cdots,n_r)$. 因此, 根据命题 6.2, 一共有 $\binom{n+r-1}{r-1}$ 项. ■

例 6d 再来讨论例 4c, 有 n 个天线, 其中 m 个是不可分辨的失效天线, 另 $n-m$ 个也是不可分辨但却有效的. 现在要求求出排成一排且没有连续两个失效天线的可能排列数. 设想 m 个失效的天线排成一排, 现找出放 $n-m$ 个有效天线的位置. 如果是排成了如下方式:

$$x_1 0 x_2 0\cdots x_m 0 x_{m+1}$$

其中 $x_1 \geqslant 0$ 是放在最左边的有效天线数; $x_i>0, i=2,\cdots,m$, 是放在第 i 个失效天线和第 $i-1$ 个失效天线之间的有效天线的个数; $x_{m+1}\geqslant 0$ 是放在最右边的有效天线数. 这样的配置意味着任两个失效天线之间都至少有一个有效天线, 因此, 满足条件的可能数是下列方程向量解的个数:

$$x_1+\cdots+x_{m+1}=n-m \qquad x_1\geqslant 0,\ x_{m+1}\geqslant 0,\ x_i>0,\ i=2,\cdots,m$$

令 $y_1=x_1+1$, $y_i=x_i$, $i=2,\cdots,m$, $y_{m+1}=x_{m+1}+1$, 可以看出它等同于以下方程的正整数向量解个数:

$$y_1+y_2+\cdots+y_{m+1}=n-m+2$$

由命题 6.1 知, 一共有 $\binom{n-m+1}{m}$ 种这样的配置方式, 这与 例 4c 的结果一致.

现在来考虑每两个失效天线之间至少有两个有效天线这种情况的排列数. 根据上述同样的理由, 结果为如下方程的解的个数:

$$x_1+\cdots+x_{m+1}=n-m \qquad x_1\geqslant 0,\ x_{m+1}\geqslant 0,\ x_i\geqslant 2,\ i=2,\cdots,m$$

令 $y_1 = x_1 + 1, y_i = x_i - 1, i = 2, \cdots, m, y_{m+1} = x_{m+1} + 1$, 可以看出它等同于以下方程的正整数向量解的个数:

$$y_1 + \cdots + y_{m+1} = n - 2m + 3$$

因此, 由命题 6.1, 可知一共有 $\binom{n-2m+2}{m}$ 种配置方式. ■

小　结

计数基本法则阐述了如下事实: 如果一个试验分成两个阶段, 第一个阶段有 n 种可能结果, 每种结果又对应于第二个阶段的 m 种可能结果, 那么该试验一共有 nm 种可能结果.

n 个元素的排列一共有 $n! = n(n-1)\cdots 3\cdot 2\cdot 1$ 种可能排列方式. 特别地, $0! = 1$.

令

$$\binom{n}{i} = \frac{n!}{(n-i)!\,i!}$$

其中 $0 \leqslant i \leqslant n$, 否则等于 0. 此式表明了从 n 个元素中选取 i 个元素的可能选取方法数, $\binom{n}{i}$ 称为从 n 个对象中选取 i 个对象的组合数, 因其在二项式定理中的突出地位, 它也常称为*二项式系数*, 我们有

$$(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i}$$

对于任意和为 n 的非负整数 $n_1, \cdots, n_r$,

$$\binom{n}{n_1, n_2, \cdots, n_r} = \frac{n!}{n_1!\,n_2!\cdots n_r!}$$

它等于 n 个元素分成互不重叠的 r 部分, 其中各个部分的元素个数分别是 n_1, $n_2, \cdots, n_r$ 的分法数.

习　题

1. (a) 前 2 位为字母而后 5 位为数字的汽车牌照一共多少种?
(b) 在上述条件下, 如果不允许字母和数字重复, 又有多少种?

2. 掷一枚骰子 4 次, 一共有多少种结果? 比如说, 如果第一次为 3, 第二次为 4, 第三次为 3, 第四次为 1, 那么结果就是“3,4,3,1”.

3. 20 种不同的工作分派给 20 个工人, 每个工人一种工作, 问有多少种分派方式?

4. 约翰、吉姆、杰伊和杰克组成一个有 4 种乐器的乐队, 如果每个人都会演奏这 4 种乐器, 问可以有多少种不同的组合? 如果约翰和吉姆会演奏这 4 种乐器, 而杰伊和杰克分别只会弹钢琴及打鼓, 那么又有多少种不同组合?

5. 一直以来, 美国和加拿大的电话区号由 3 位数组成, 第一位是 $2 \sim 9$ 之间的任一数字; 第二位是 0 或者 1; 第三位是 $1 \sim 9$ 之间的任一数字. 问一共有多少种可能的区号? 以 4 开头的区号一共有多少种可能?

6. 有个熟知的童谣:
我出发去圣 - 艾弗斯,
路上碰见一个人,
带着他的 7 个老婆,
每个老婆挎着 7 个包,
每个包里装着 7 只猫,
每只猫生了 7 只小猫咪,
数数看, 我到底看到了多少只小猫咪?

7. (a) 3 个男孩和 3 个女孩坐在一排, 一共有几种坐法?
(b) 如果男孩和女孩分别坐在一起, 一共有几种坐法?
(c) 如果只要求男孩坐在一起, 一共有几种坐法?
(d) 如果相邻的座位上必须坐异性, 一共有几种坐法?

8. 以下单词的字母可以有多少种排列方式?
(a) Fluke; (b) Propose; (c) Mississippi; (d)Arrange.

9. 一个孩子有 12 块积木, 其中 6 块黑色、4 块红色、1 块白色和 1 块蓝色. 孩子想把这些积木排成一排, 一共有多少种排法?

10. 8 人坐在一排, 一共有多少种坐法, 如果
(a) 没什么限制; (b) A 和 B 必须坐在一起; (c) 一共 4 个男人, 4 个女人, 且任何两个男人不能坐在一起, 任何两个女人也不能坐在一起; (d) 共有 5 个男人, 且他们必须坐在一起; (e) 有 4 对夫妇, 每对夫妇必须坐在一起.

11. 把 3 本小说、2 本数学书和 1 本化学书摆放到书柜里, 一共有多少种摆放方法? 如果
(a) 书可以以任何顺序摆放; (b) 数学书必须放一起, 小说必须放一起; (c) 小说必须放一起, 其他书无所谓.

12. 某个班级有 30 名学生, 有 5 个不同的奖项 (成绩奖、组织奖等等) 要颁发, 一共有多少种不同的颁奖方式? 如果
(a) 一个学生可以得多个奖项; (b) 每个学生最多只能得 1 个奖项.

13. 有 20 个人, 每人都要与其他人握一次手, 一共要握多少次手?

14. 从一副牌的 52 张中任意抽取 5 张, 一共有多少种抽法?

15. 一个舞蹈班有 22 个学生, 10 女 12 男, 要挑选 5 男 5 女然后配对, 一共有多少种配法?

16. 某学生从 6 本数学书、7 本科学书和 4 本经济学书里卖掉 2 本, 有多少种选择? 如果
(a) 两本书同一科目; (b) 两本书不同科目.

17. 共有 7 件不同的礼物分给 10 个孩子, 如果每个孩子最多只能拿 1 件礼物, 一共有多少种分法?

18. 有 5 名共和党人, 6 名民主党人和 4 名无党派人士, 要从中分别选取 2 名共和党人、2 名民主党员和 3 名无党派人士组成一个七人委员会. 一共有多少种选法?

19. 从 8 女 6 男里选择 3 女 3 男组成委员会, 一共有多少种选法? 如果

(a) 其中有两个男人不愿同时进入委员会;

(b) 其中有两个女人不愿同时进入委员会;

(c) 其中有一男一女不愿同时进入委员会.

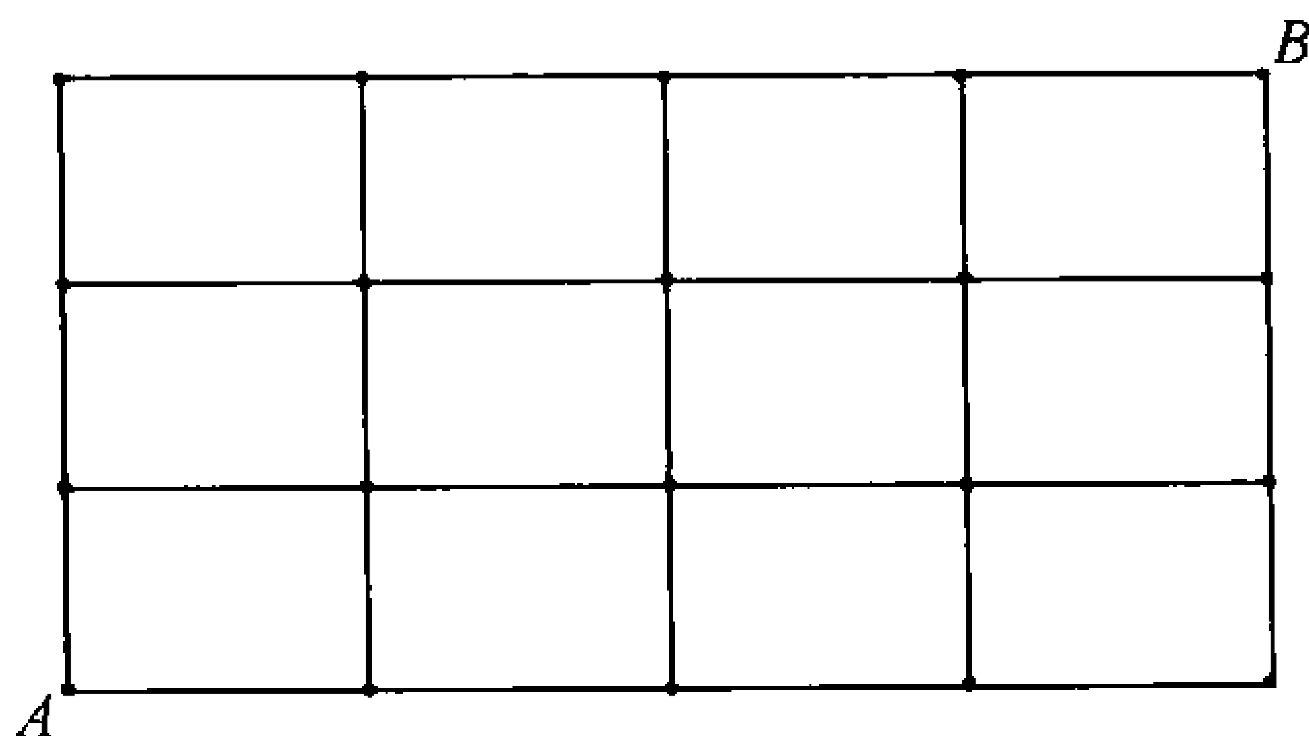
20. 某人有 8 个朋友, 他打算邀请其中 5 人参加茶会,

(a) 如果有某两人长期不和, 不能同时参加茶会, 他有多少种邀请方案?

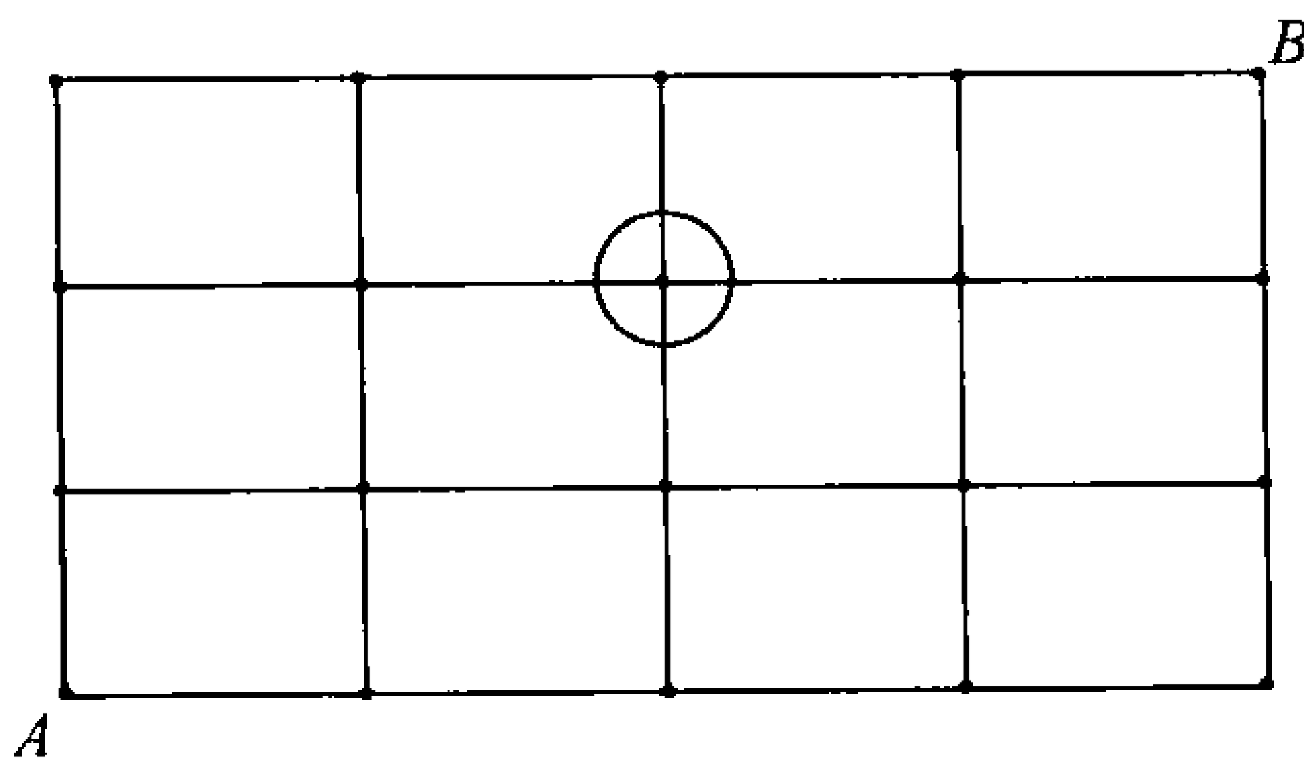
(b) 如果有某两人只能同时被邀请, 一共有多少种邀请方案?

21. 如下图, 从标有 A 的地方出发, 每一步只能向上或者向右移动, 问移动到 B 一共有多少种移动方式?

提示: 从 A 到 B 需向右 4 步, 向上 3 步.



22. 在习题 21 中, 如果要求必须经过标有圆圈的点 (如下图), 一共有多少种移动方式?



23. 某从事睡梦研究的心理学实验室有 3 间房, 每间房里有 2 张床. 现要安排 3 对双胞胎进行睡梦实验, 要求每对双胞胎必须安排在同一间房的不同床上, 一共有多少安排方法?

24. 展开 $(3x^2+y)^5$.

25. 桥牌比赛中有四个选手参加, 每人分到 13 张牌, 一共有多少种分法?

26. 展开 $(x_1+2x_2+3x_3)^4$.

27. 将 12 个人分成各含 3 人、4 人和 5 人的委员会, 一共有多少种方法?

28. 将 8 个老师分配到 4 个学校, 一共有多少种分法? 如果要求每个学校必须接收 2 个老师呢?

29. 一共 10 个举重选手参加比赛, 其中 3 名美国选手、4 名俄罗斯选手、2 名中国选手和 1 名加拿大选手. 如果成绩只记录每个选手的国籍, 一共有多少种可能结果? 如果美国选手有 1 名在总成绩前三名, 另两名在总成绩的最后三名中, 那么一共有多少种可能结果?

30. 有分别来自俄国、法国、英国和美国等 10 个国家的代表坐在一排, 如果法国和英国代表坐在一起, 俄国和美国代表不坐在一起, 一共有多少种坐法?

***31.** 8 块相同的黑板分给 4 所学校, 一共有多少种分法? 如果要求每所学校至少分到 1 块黑板呢?

***32.** 电梯载着 8 个人 (不包括电梯工) 自底层启动, 到顶层 6 楼后, 乘客已全部下完. 如果电梯工只注意到每层楼出去的人数, 那么他能看到多少种离开电梯的方式? 如果 8 个乘客中有 5 个男人和 3 个女人, 而电梯工又只注意了出去人的性别; 问题的答案又是多少?

***33.** 有 2 万美元要投资到 4 个项目上, 每份投资必须是 1000 美元的整数倍, 且每个项目如果有投资的话, 最少投资额分别为 2000, 2000, 3000 和 4000 美元, 一共有多少种可行的投资方法, 如果

(a) 每个项目都要投资; (b) 至少投资其中 3 个项目.

理论习题

1. 证明推广的计数基本法则.

2. 进行两个试验, 第一个试验有 m 个可能的结果. 若第一个试验得到第 i 个结果, 则第二个试验有 n_i 个可能的结果, $i=1,2,\cdots,m$. 问这两个试验一共有多少种可能结果?

3. 从 n 个元素里取 r 个, 如考虑抽取次序的话有多少种取法?

4. 有 n 个球, 其中 r 个黑球, $n-r$ 个白球, 把它们排成一排, 用组合学知识解释共有 $\binom{n}{r}$ 种排法.

5. 计算形如 $(x_1,x_2,\cdots,x_n)$ 的向量的个数, 其中 x_i 等于 0 或者 1, 且 $\sum_{i=1}^{n} x_i \geqslant k$

6. 有多少个这样的向量 $(x_1,\cdots,x_k)$, 其中 x_i 是正整数, 且满足 $1\leqslant x_i\leqslant n$ 和 $x_1<x_2<\cdots<x_k$?

7. 用分析的方法证明等式 (4.1).

8. 证明 $\binom{n+m}{r}=\binom{n}{0}\binom{m}{r}+\binom{n}{1}\binom{m}{r-1}+\cdots+\binom{n}{r}\binom{m}{0}$

提示: 设有 n 个男人和 m 个女人, 从中挑选 r 人, 一共有多少种挑选方法?

9. 利用理论习题 8 的结论证明: $\binom{2n}{n}=\sum_{k=0}^{n}\binom{n}{k}^2$

10. 从 n 人中选取 k 人组成一个委员会, $k\leqslant n$, 其中一人被任命为主席.

(a) 考虑先选出 k 人, 然后任命其中一人为主席, 说明总共有 $\binom{n}{k}k$ 种可能方式.

(b) 考虑先选出 $k-1$ 人, 其中没有主席, 然后再在剩下的 $n-k+1$ 人中选一人为主席, 总共有 $\binom{n}{k-1}(n-k+1)$ 种可能方式.

(c) 考虑先选出主席, 然后再选出其他成员, 说明一共有 $n\binom{n-1}{k-1}$ 种可能选择.

(d) 总结 (a), (b), (c), 得出 $k\binom{n}{k}=(n-k+1)\binom{n}{k-1}=n\binom{n-1}{k-1}$

(e) 利用 $\binom{m}{r}$ 的阶乘定义证明 (d) 中等式.

11. 以下是费马组合恒等式:

$$\binom{n}{k}=\sum_{i=k}^{n}\binom{i-1}{k-1} \qquad n \geqslant k$$

试从组合的角度, 而不用计算的方法, 去验证该恒等式.

提示: 考虑由数字 $1,\cdots,n$ 所组成的集合, 以 i 为最大值的含有 k 个元素的子集一共有多少个?

12. 考虑如下组合恒等式:

$$\sum_{k=1}^{n} k\binom{n}{k}=n\cdot 2^{n-1}$$

(a) 试从组合的角度解释上式, 可考虑从 n 人中挑选若干人组成委员会并在其中选定一名主席的可能方式的两种计算方法.

提示:

(i) 如果选择 k 人组成会员会并选定一名主席, 一共有多少种方法?

(ii) 先选好主席, 然后再选其他成员, 一共有多少种选法?

(b) 证明以下等式对 $n=1,2,3,4,5$ 都成立:

$$\sum_{k=1}^{n}\binom{n}{k}k^2=2^{n-2}n(n+1)$$

为了从组合的角度证明上式, 指出上式的两边都等于如下的可能选择方式: 考虑有 n 个人, 从中选择若干人组成一个委员会, 并选定主席和秘书 (有可能是同一人).

提示:

(i) 委员会一共有 k 人, 一共有多少种选择方式?

(ii) 如果主席和秘书为同一人, 一共有多少种选择方式? (答案: $n2^{n-1}$)

(iii) 如果主席和秘书为两个不同的人, 一共有多少种选择方式?

(c) 证明 $\sum_{k=1}^{n}\binom{n}{k}k^3=2^{n-3}n^2(n+3)$

13. 证明, 对 $n>0$, 有 $\sum_{i=0}^{n}(-1)^i\binom{n}{i}=0$

提示: 利用二项式定理.

14. 从 n 个人里选 j 个人组成委员会, 再从这委员会里选 i 个人组成分会, $i \leqslant j$.

(a) 用两种方法分别计算委员会和分会的可能选择数来导出组合恒等式. 其中, 第一种方法是先选择 j 个人组成委员会, 再从中选择 i 个人组成分会. 第二种是先选择 i 个人组成分会, 再补充 $j-i$ 个人组成委员会.

(b) 利用 (a) 证明组合恒等式: $\sum_{j=1}^{n}\binom{n}{j}\binom{j}{i}=\binom{n}{i}2^{n-i} \qquad i \leqslant n$

(c) 利用 (a) 和理论习题 13 证明 $\sum_{j=1}^{n}\binom{n}{j}\binom{j}{i}(-1)^{n-j}=0 \qquad i \leqslant n$

15. 令 $H_k(n)$ 为向量 $(x_1,x_2,\cdots,x_k)$ 的数目, 其中 x_i 是正整数且满足 $1 \leqslant x_i \leqslant n$ 及 $x_1 \leqslant x_2 \leqslant \cdots \leqslant x_k$.

(a) 不用任何计算, 说明

$$H_1(n) = n$$
$$H_k(n) = \sum_{j=1}^{n} H_{k-1}(j) \qquad k > 1$$

提示: 如果 $x_k = j$, 那一共有多少向量?

(b) 利用上述递推公式计算 $H_3(5)$.

提示: 先计算 $H_2(n)$, 对 $n = 1, 2, 3, 4, 5$.

16. 有 n 个选手参加比赛, 最后排定成绩, 同时允许选手排名相同. 那么, 就可按排名成绩将选手分成组, 成绩最好的第一组, 成绩其次的第二组, 等等. 记 $N(n)$ 表示不同结果的可能数, 比如 $N(2) = 3$, 因为在一个只有 2 名选手参加的比赛中, 比赛结果一共有 3 种: 第一个选手获第一, 第二个选手获第一, 两个选手并列第一.

(a) 列出所有 $n = 3$ 时的可能结果;

(b) 令 $N(0) = 1$, 不用任何计算, 说明 $N(n) = \sum_{i=1}^{n} \binom{n}{i} N(n-i)$ **提示**: 共有 i 个选手并列最后一名, 一共有多少种结果?

(c) 证明上述公式等价于 $N(n) = \sum_{i=0}^{n-1} \binom{n}{i} N(i)$

(d) 利用上述递推公式, 求出 $N(3)$ 和 $N(4)$.

17. 从组合的角度解释 $\binom{n}{r} = \binom{n}{r, n-r}$.

18. 证明

$$\binom{n}{n_1, n_2, \cdots, n_r} = \binom{n-1}{n_1 - 1, n_2, \cdots, n_r} + \binom{n-1}{n_1, n_2 - 1, \cdots, n_r} + \cdots + \binom{n-1}{n_1, n_2, \cdots, n_r - 1}$$

提示: 利用证明公式 (4.1) 的类似方法.

19. 证明多项式定理.

***20.** 将 n 个相同的球放到 r 个坛子里, 要求第 i 个坛子至少有 m_i 个球, $i = 1, \cdots, r$, 一共有多少种放法? 假设 $n \geqslant \sum_{i=1}^{r} m_i$.

***21.** 说明方程 $x_1 + x_2 + \cdots + x_r = n$ 正好有 $\binom{r}{k}\binom{n-1}{n-r+k}$ 个解, 其中有 k 个 x_i 为 0.

***22.** 考虑 n 元函数 $f(x_1, \cdots, x_n)$, 一共有多少个 r 阶偏微分?

***23.** 求向量 $(x_1, \cdots, x_n)$ 的数目, 其中 x_i 为非负整数且满足 $\sum_{i=1}^{n} x_i \leqslant k$.

自 检 习 题

1. 字母 A, B, C, D, E,F 一共有多少种排列方式, 如果

(a) A 和 B 必须在一起;

(b) A 在 B 之前;

(c) A 在 B 之前, B 在 C 之前;

(d) A 在 B 之前, C 在 D 之前;

(e) A 和 B 必须在一起, C 和 D 也必须在一起;

(f) E 不在最后.

2. 4 个美国人、3 个法国人和 3 个英国人坐在一排, 要求相同国籍的人必须坐在一起, 一共有多少种坐法?

3. 从有 10 人的俱乐部中分别选 1 名总裁、1 名财务和 1 名秘书, 一共有多少种选法, 如果

(a) 没有任何限制; (b) A 和 B 不能同时被选;

(c) C 和 D 要么同时被选, 要么同时不被选; (d) E 必须被选;

(e) F 被选中的话, 必须担任总裁.

4. 一次考试中, 学生应从 10 道考题中选择 7 道回答, 一共有多少种选法? 如果规定必须在前 5 道中至少选 3 道, 那么有多少种选法?

5. 某人将 7 件礼物分给他的 3 个孩子, 其中老大得 3 件, 其余两人分别得 2 件, 一共有多少分法?

6. 一块汽车牌照的号码由 3 位字母和 4 位数字组成. 如果允许字母或数字重复且位置没有任何限制, 一共有多少种可能的牌照号?

7. 从组合的角度解释恒等式: $\binom{n}{r} = \binom{n}{n-r}$.

8. 考虑一个 n 位数, 每位数字都是 $0, 1, \cdots, 9$ 中的一个, 一共有多少个这样的数? 如果

(a) 相邻的两位数字均不相同; (b) 0 出现 i 次, $i = 0, \cdots, n$.

9. 有三个班级, 每个班级有 n 个学生, 从这 $3n$ 个学生中选 3 人.

(a) 一共有多少种选法?

(b) 如果 3 人来自于相同班级, 一共有多少种选法?

(c) 如果有 2 人来自于相同班级, 另 1 人来自于其他班级, 一共有多少种选法?

(d) 3 人分别来自不同的班级, 一共有多少种选法?

(e) 利用 (a) 到 (d) 的结果, 写出一个组合恒等式.

10. 由数字 $1, 2, \cdots, 9$ 组成的 5 位数一共有多少? 这 5 个数字中不容许有数字重复多于 2 次 (例如, 41 434 是不容许的).

11. 从 10 对夫妇中选出 6 个人, 但不容许夫妇同时入选.

(a) 一共有多少种不同的选法?

(b) 若规定这 6 个人中必须是三男三女 (当然夫妇是不容许同时入选的), 一共有多少种选法?

12. 从 7 个男人、8 个女人中选取 6 人组成委员会. 如果要求至少 3 个女人、2 个男人, 一共有多少种选取方法?

***13.** 一个艺术品收藏拍卖会一共有 15 件艺术品, 其中 4 件达利的、5 件凡高的和 6 件毕加索的. 一共有 5 名艺术品收藏家买下了这批艺术品. 而某记者只记载了每位收藏家得到的达利、凡高和毕加索作品的数量, 问销售记录能有多少种不同的结果?

***14.** 计算向量 $(x_1, \cdots, x_n)$ 的个数, 如果 x_i 是正整数, 且 $\sum_{i=1}^{n} x_i \leqslant k$, 其中 $k \geqslant n$.

15. n 个学生参加保险精算师的考试, 公榜结果只列出那些通过考试的学生名单, 并且按照他们的分数由高到底进行排序, 例如, 公榜结果为"Brown, Cho"意味着只有 Brown 和 Cho 通过了考试, 而且 Brown 分数比 Cho 的高. 如果没有相同的分数, 那么公布的考试结果一共有多少种情况?

16. 从集合 $S = \{1, 2, \cdots, 20\}$ 中选 4 个元素组成子集, 并且 $1, 2, 3, 4, 5$ 中至少有一个被选中, 一共有多少种子集?

17. 给出下列恒等式分析的证明

$$\binom{n}{2}=\binom{k}{2}+k(n-k)+\binom{n-k}{2} \qquad 1\leqslant k\leqslant n$$

并给出该恒等式的组合解释.

18. 一个社区的构成是这样的: 有 3 户单亲家庭带 1 个子女, 有 3 户单亲家庭带 2 个子女, 有 5 户双亲家庭带一独生子女, 7 户双亲家庭带 2 个子女, 6 户双亲家庭带了 3 个子女. 现在要从某一家庭中选取一个家长和一个孩子, 一共有多少种选法?

19. 汽车牌照号由 8 位数字或字母组成, 但规定数字或字母不能重复, 对于数字或字母的位置不作任何限制, 这样共有多少种不同的汽车牌号? 若还规定 3 个数字必须是相连的, 那么问一共可有多少种不同的汽车牌号?

第 2 章　概率论公理化

2.1　简　介

本章介绍事件的概率这一概念, 并展示在某些特定情形下计算概率的方法. 在引入概率的概念之前, 先要学习更基本的概念 —— 试验的样本空间和事件.

2.2　样本空间和事件

假设某次试验的结果是不可预测、不确定的. 当然, 尽管在试验之前无法得知结果, 但是假设所有可能的结果的集合是知道的. 所有可能的结果构成的集合, 称为该试验的**样本空间**(sample space), 并记为 S. 以下是样本空间的一些例子.

(1) 若试验是考察新生婴儿的性别, 那么所有可能结果的集合 $S=\{\mathrm{g},\mathrm{b}\}$ 就是一个样本空间, 其中 g 表示“女孩”, b 表示“男孩”.

(2) 赛马比赛中一共有 7 匹马参赛, 这 7 匹马分别标以 $1,2,3,4,5,6,7$. 考察比赛结果, 所有可能的比赛结果的集合 $S=\{(1,2,3,4,5,6,7)$的所有$7!$种排列$\}$ 就是一个样本空间. 比如 $(2,3,1,6,5,4,7)$ 就表示 2 号马跑第一, 3 号马跑第二, 接下来是 1 号马……这是比赛的一种可能结果.

(3) 掷两枚硬币, 考察哪一面朝上, 那么样本空间一共包含如下四种结果:

$$S=\{(\mathrm{H},\mathrm{H}),(\mathrm{H},\mathrm{T}),(\mathrm{T},\mathrm{H}),(\mathrm{T},\mathrm{T})\}$$

其中 (H,H) 表示两枚硬币都是正面朝上[①]; (H,T)表示第一枚硬币正面朝上, 第二枚反面朝上; (T,H)表示第一枚硬币反面朝上, 第二枚正面朝上; (T,T)表示两枚硬币都是反面朝上.

(4) 掷两枚骰子, 考察两枚骰子的点数, 那么样本空间包含 36 个结果:

$$S=\{(i,j):i,j=1,2,3,4,5,6\}$$

其中, (i,j) 表示第一个骰子的点数是 i, 第二个骰子的点数是 j.

(5) 考察一个晶体管的寿命 (小时), 那么样本空间是所有的非负实数, 即

$$S=\{x:0\leqslant x<\infty\}$$

① 本书中, 我们约定 H 表示正面朝上, T 表示反面朝上.—— 译者注

样本空间的任一子集 E 称为**事件**(event), 事件就是由试验的某些可能结果组成的一个集合. 如果试验的结果包含在 E 里面, 那么就称 E 发生了. 以下是一些有关事件的例子:

在前面的例 (1) 中, 令 $E = \{g\}$, 那么 E 就表示"婴儿是个女孩"这个事件; 类似地, $F = \{b\}$ 就表示"婴儿是个男孩".

在例 (2) 中, 如果 $E=\{$所有以 3 开头的排列$\}$, 那么 E 就表示"3 号马获得了第一".

在例 (3) 中, 如果 $E = \{(H,H),(H,T)\}$, 那么 E 就表示"第一枚硬币正面朝上".

在例 (4) 中, 如果 $E = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$, 那么 E 就表示"两个骰子点数之和为 7".

在例 (5) 中, 如果 $E = \{x : 0 \leqslant x \leqslant 5\}$, 那么 E 就表示"晶体管的寿命不超过 5 个小时".

对于同一个样本空间 S 的任两个事件 E 和 F, 定义一个新的事件 $E \cup F$, 它由以下结果组成: 这些结果或在 E 里, 或在 F 里, 或既在 E 里也在 F 里. 也就是说, 如果事件 E 或者事件 F 中有一个发生, 那么 $E \cup F$ 就发生. 比如在例 (1) 中, 如果 $E = \{g\}, F = \{b\}$, 那么 $E \cup F = \{g,b\}$, 也即, $E \cup F$ 与整个样本空间 S 是一致的. 在例 (3) 中, 如果 $E = \{(H,H),(H,T)\}$ 且 $F = \{(T,H)\}$, 那么 $E \cup F = \{(H,H),(H,T),(T,H)\}$, 因此, $E \cup F$ 意味着"至少有一枚硬币正面朝上".

事件 $E \cup F$ 称为事件 E 和事件 F 的**并**(union).

类似地, 对于任意两个事件 E 和 F, 还可以定义 EF, 称为 E 和 F 的**交**(intersection), 它由 E 和 F 的公共元素组成. 也即事件 EF(有时也记为 $E \cap F$) 发生当且仅当 E 和 F 同时发生. 比如在例 (3) 中, 事件 $E = \{(H,H),(H,T),(T,H)\}$ 表示"至少有一个正面朝上", 而 $F = \{(H,T),(T,H),(T,T)\}$ 表示"至少有一个反面朝上", 那么 $EF = \{(H,T),(T,H)\}$ 就表示"正好一个正面朝上, 一个反面朝上". 在例 (4) 中, 事件 $E = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$ 表示两个骰子点数之和为 7, 而 $F = \{(1,5),(2,4),(3,3),(4,2),(5,1)\}$ 表示骰子点数之和为 6, 那么 EF 不包含任何试验结果, 所以它也不可能发生. 类似这样的事件, 称为不可能事件, 记为 $\varnothing$(也即, $\varnothing$ 是不包含任何结果的事件). 如果 $EF = \varnothing$, 则称 E 和 F 是**互不相容的**(mutually exclusive).

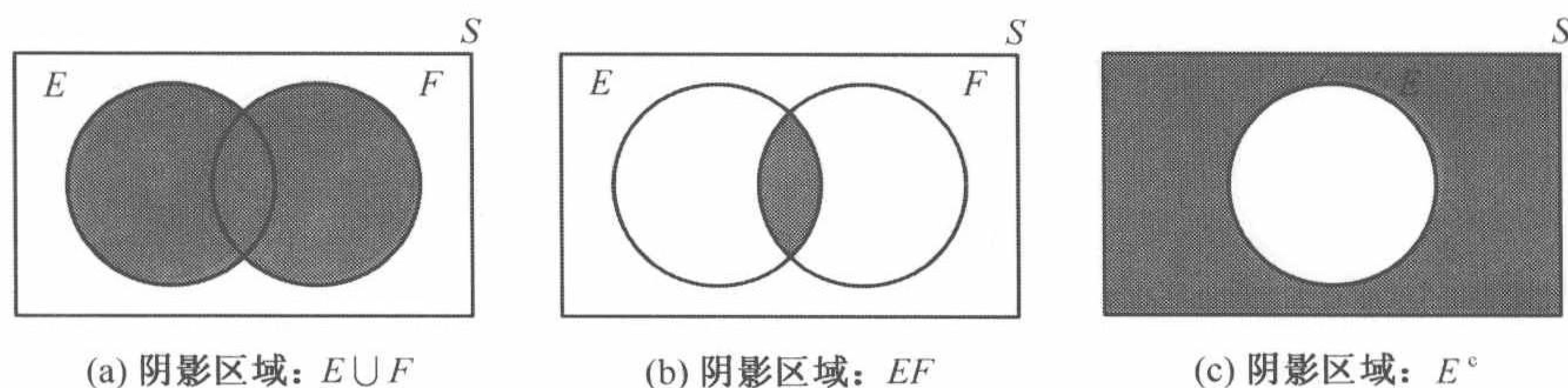
用类似的方式再来定义两个以上事件的并和交. 设 $E_1, E_2, \cdots$ 是一系列事件, 这些事件的并记为 $\bigcup_{n=1}^{\infty} E_n$, 表示至少包含在某一个 E_n 里的所有结果所构成的事件. 同样, 这些事件的交记为 $\bigcap_{n=1}^{\infty} E_n$, 其含义为包含在所有 E_n 里的所有结果构成的事件.

最后, 一个事件的补事件记为 E^c, 其含义是包含在样本空间里但不包含在 E 里的所有结果构成的事件. 也即, E^c 发生当且仅当 E 不发生. 在例 (4) 中, 如

$E=\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$, 那么当两个骰子点数之和不等于 7 时, E^c 发生. 值得说明的是, 样本空间 S 的补集 $S^c=\varnothing$.

对于任两个事件 E 和 F, 如果 E 内的所有结果都在 F 中, 那么称 E 包含于 F, 记为 $E\subset F$(或者 $F\supset E$). 因此, 如果 $E\subset F$, 那么 E 发生也就能推出 F 也发生. 如果 $E\subset F$ 和 $F\subset E$, 那么称 E 和 F 是相同的, 记为 $E=F$.

韦恩图 (Venn Diagram)是一种用来阐述事件之间的逻辑关系的非常有效的几何表示方法. 样本空间 S 表示为一个大的矩形, 表示包含了所有可能结果, 事件 $E,F,G,\cdots$ 表示为包含在矩形之内的一个个小圆, 所关心的事件可以用相应的阴影区域来表示. 比如, 在图 2.1 的三个韦恩图中, 阴影部分分别表示 $E\cup F, EF$ 和 E^c. 图 2.2 表示 $E\subset F$.



(a) 阴影区域: $E\cup F$　(b) 阴影区域: EF　(c) 阴影区域: E^c

图 2.1 韦恩图

并、交和对立事件遵循类似于代数学里的一些运算规则, 列举如下:

交换律: $E\cup F=F\cup E\ EF=FE$

结合律: $(E\cup F)\cup G=E\cup(F\cup G)$
$(EF)G=E(FG)$

分配律: $(E\cup F)G=EG\cup FG\ EF\cup G=(E\cup G)(F\cup G)$

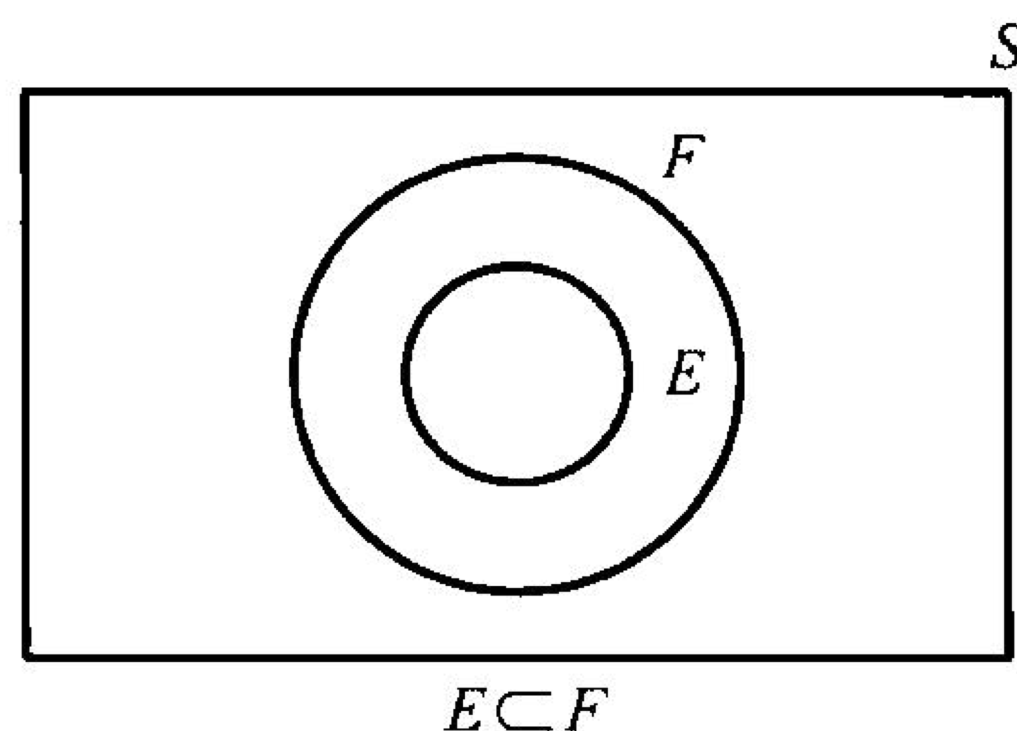
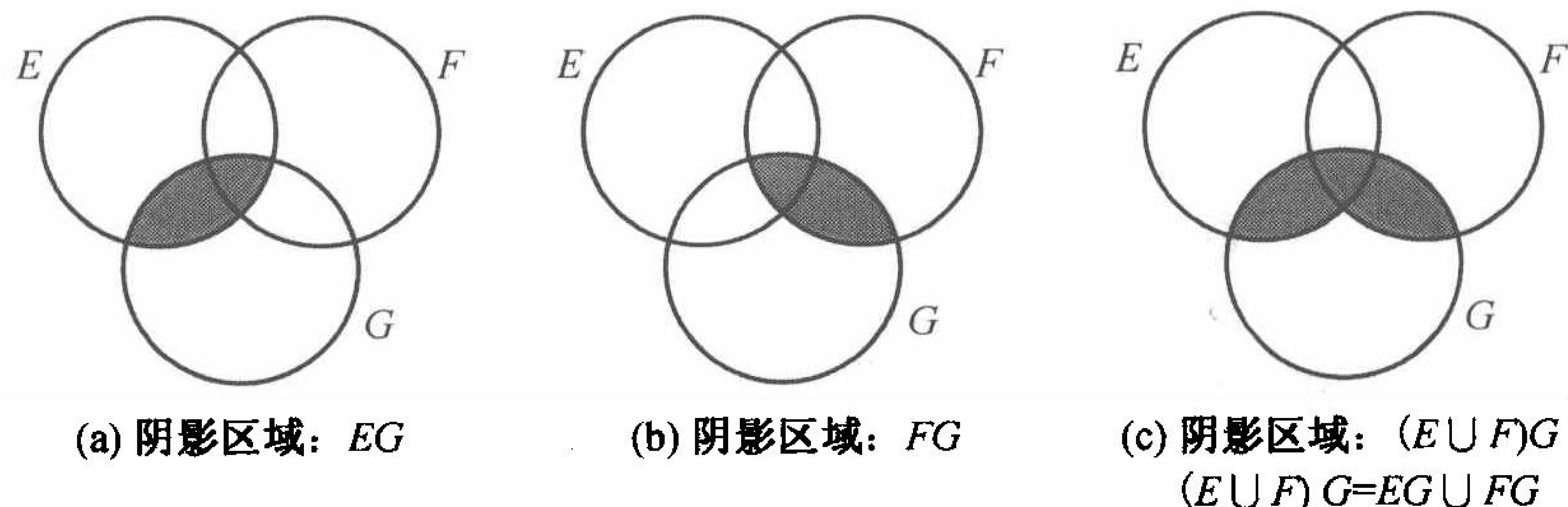


图 2.2 事件的包含关系

上述式子可以通过以下方式来证明: 先证明等式左边的事件里的任一结果必然包含于等式右边的事件, 再证明等式右边的事件里的任一结果也包含于等式左边的事件. 另一个显而易见的方法就是利用韦恩图, 比如图 2.3 就表示了分配律.



(a) 阴影区域: EG　(b) 阴影区域: FG　(c) 阴影区域: $(E\cup F)G$
$(E\cup F)G=EG\cup FG$

图 2.3 分配律

下面关于并、交及对立事件这三个基本运算之间的重要的关系式称为德摩根定律 (DeMorgan's laws)：

$$\left(\bigcup_{i=1}^{n} E_i\right)^{\mathrm{c}} = \bigcap_{i=1}^{n} E_i^{\mathrm{c}} \quad \left(\bigcap_{i=1}^{n} E_i\right)^{\mathrm{c}} = \bigcup_{i=1}^{n} E_i^{\mathrm{c}}$$

为了证明上述定律, 首先假设 x 是 $\left(\bigcup_{i=1}^{n} E_i\right)^{\mathrm{c}}$ 里的一个元素, 那么 x 不包含于 $\bigcup_{i=1}^{n} E_i$, 这就意味着 x 并不包含于任一个事件 $E_i, i = 1, 2, \cdots, n$, 所以对任意 $i(i = 1, 2, \cdots, n)$ 来说, x 就包含于 E_i^{c}, 也即 x 包含于 $\bigcap_{i=1}^{n} E_i^{\mathrm{c}}$. 另一方面, 假设 x 包含于 $\bigcap_{i=1}^{n} E_i^{\mathrm{c}}$, 那么对任一 i, $i = 1, 2, \cdots, n$, x 包含于 E_i^{c}. 这就意味着 x 不属于所有的 E_i, 即 x 不包含于 $\bigcup_{i=1}^{n} E_i$. 也即 x 包含于 $\left(\bigcup_{i=1}^{n} E_i\right)^{\mathrm{c}}$. 这样就证明了德摩根定律的第一条.

现证明德摩根定律的第二条, 由第一条定律可知

$$\left(\bigcup_{i=1}^{n} E_i^{\mathrm{c}}\right)^{\mathrm{c}} = \bigcap_{i=1}^{n} (E_i^{\mathrm{c}})^{\mathrm{c}}$$

这样, 由 $(E^{\mathrm{c}})^{\mathrm{c}} = E$, 上式等价于

$$\left(\bigcup_{i=1}^{n} E_i^{\mathrm{c}}\right)^{\mathrm{c}} = \bigcap_{i=1}^{n} E_i$$

对两边取补运算, 即得到如下结果：

$$\bigcup_{i=1}^{n} E_i^{\mathrm{c}} = \left(\bigcap_{i=1}^{n} E_i\right)^{\mathrm{c}}$$

2.3　概率论公理

一种定义事件发生的概率的方式是利用事件发生的频率. 定义如下：一个试验的样本空间为 S, 在相同的条件下可重复进行. 对于样本空间 S 里的事件 E, 记 $n(E)$ 为 n 次重复试验中事件 E 发生的次数. 那么, 该事件发生的概率

$$P(E) = \lim_{n \to \infty} \frac{n(E)}{n}$$

即概率 $P(E)$ 定义为 E 发生的次数占试验总次数的比例的极限, 也即发生频率的极限.

虽然上述定义很直观, 而且大多读者也一直这么认为, 但它却有很严重的缺陷. 怎么就知道 $n(E)/n$ 会收敛到一个固定的常数, 而且如果进行另一次重复试验, 它也会收敛到这个相同的常数? 例如, 设想进行这样的试验：重复掷一枚硬币 n 次, 怎么能保证在 n 次试验中正面朝上的比例会随着 n 的增大而收敛于某个数? 而且, 即使它确实收敛于某个数, 又如何保证进行另一次同样的重复试验时, 其比例会趋

于同样的值?

用频率来定义概率的支持者常常这样回答上述问题, 他们认为 $n(E)/n$ 趋于某常数是整个系统的一个假设, 或者说一个**公理**(axiom). 但是, 这个假设非常不简洁. 因为, 尽管事实上需要假定频率的极限是存在的, 但是这却不是一个最基本、最简单的假设. 同时, 这样的假设也不一定为所有人所认同. 事实上, 先假定一些更简单、更显而易见的关于概率的公理, 然后去证明频率在某种意义下趋于一个常数极限不是更合情合理吗? 这也正是本书采纳的现代概率论公理化方法. 特别地, 仍假定对于样本空间里的任一事件 E, 都存在一个值 $P(E)$(指的就是事件 E 的概率), 并假定这些概率值符合一系列公理. 读者一定会认可这些公理, 因为这些公理很接近于对概率的直觉认识.

假设某个试验的样本空间为 S, 对应于其中任一事件 E, 定义一个数 $P(E)$, 满足如下 3 条公理.

公理 1 $0 \leqslant P(E) \leqslant 1$

公理 2 $P(S)=1$

公理 3 对任一系列互不相容的事件 $E_1, E_2, \cdots$(即如果 $i \neq j$, 则 $E_iE_j=\varnothing$), 有

$$P\left(\bigcup_{i=1}^{\infty} E_i\right)=\sum_{i=1}^{\infty} P(E_i)$$

我们把满足以上 3 条公理的 $P(E)$ 称为事件 E 的概率.

公理 1 说明, 任何事件 E 的概率在 0 到 1 之间. 公理 2 说明, S 作为必然发生的事件, 其概率定义为 1. 公理 3 说明对任意一列互不相容事件, 至少有一事件发生的概率等于各事件发生的概率之和. 这些公理简明又直观.

设 $E_1, E_2, \cdots$ 为一特殊的事件序列, 其中 $E_1=S, E_i=\varnothing, i>1$, 此时各个事件互不相容, 且 $S=\bigcup_{i=1}^{\infty} E_i$. 由公理 3 可以得到: $P(S)=\sum_{i=1}^{\infty} P(E_i)=P(S)+\sum_{i=2}^{\infty} P(\varnothing)$ 这就说明 $P(\varnothing)=0$, 也即空事件发生的概率为 0.

值得注意的是, 对于有限个互不相容事件的序列 $E_1, E_2, \cdots, E_n$, 有

$$P\left(\bigcup_{i=1}^{n} E_i\right)=\sum_{i=1}^{n} P(E_i) \tag{3.1}$$

为证明这个结论, 只需在公理 3 中, 令所有 $E_i(i>n)$ 为空事件即可. 当样本空间为有限集时, 公理 3 与上式是等价的, 但当样本空间是无限集时, 公理 3 的关于事件概率可加性的推广就是必要的.

例 3a 掷一枚硬币, 记正面朝上的事件为 H, 反面朝上的事件为 T. 假设两者发生的可能性一样, 那么 $P(\{\mathrm{H}\})=P(\{\mathrm{T}\})=\frac{1}{2}$.

另外, 如果这个硬币有偏向, 而且正面朝上的机会是反面朝上机会的 2 倍, 那么 $P(\{\text{H}\}) = \frac{2}{3}, P(\{\text{T}\}) = \frac{1}{3}$. ■

例 3b 掷一枚骰子, 若 6 个面出现的可能性是一样的, 这样就有 $P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\}) = 1/6$. 从公理 3 可知出现偶数面朝上的概率为 $P(\{2,4,6\}) = P(\{2\}) + P(\{4\}) + P(\{6\}) = \frac{1}{2}$. ■

设 $P(E)$ 是定义在样本空间里的事件的集函数, 若它满足公理 1, 2, 3, 则 $P(E)$ 就是事件 E 的概率. 这一定义是现代概率论的数学基础. 我们认为, 读者会认为这些公理很自然, 而且与对概率的直觉概念 (也即概率是与机会和随机性有关的知识) 很吻合. 进一步, 利用这些公理可以证明, 随着试验的不断进行, 事件 E 发生的频率趋近 $P(E)$ 的概率为 1. 这就是第 8 章将要介绍的强大数定律. 另外, 2.7 节还将介绍概率的另一种解释, 即概率可作为确信程度的度量.

技术注释 我们假定概率 $P(E)$ 是针对样本空间里的所有事件 E 定义的, 事实上, 如果样本空间是不可数集, 那么 $P(E)$ 仅仅针对那些所谓可测的事件进行定义. 但是, 这并不是概率论的缺陷, 因为所有现实中的事件都是可测的.

2.4 几个简单命题

这一节证明几个有关概率的简单性质. 首先注意到 E 和 E^c 总是互不相容的, 而且 $E \cup E^c = S$, 因此由公理 2 和公理 3 可以得到

$$1 = P(S) = P(E \cup E^c) = P(E) + P(E^c)$$

以下的命题 4.1 便是上式的等价形式.

命题 4.1 $P(E^c) = 1 - P(E)$

命题 4.1 说明, 一个事件不发生的概率, 等于 1 减去它发生的概率. 比如, 掷一枚硬币, 若正面朝上的概率是 3/8, 那么反面朝上的概率一定是 5/8.

接下来的第二个命题指出了如果事件 E 包含于事件 F, 那么 E 发生的概率必然不大于 F 发生概率.

命题 4.2 如果 $E \subset F$, 那么 $P(E) \leqslant P(F)$.

证明: 由 $E \subset F$, 可将 F 表示为 $F = E \cup E^cF$, 这样, 因为 E 和 E^cF 是互不相容的, 由公理 3 可得 $P(F) = P(E) + P(E^cF)$. 由于 $P(E^cF) \geqslant 0$, 因此 $P(E) \leqslant P(F)$. □

举个例子, 掷一枚骰子, 出现 1 的概率肯定小于等于出现奇数的概率.

下一命题借助两事件的概率给出了它们的并的概率与交的概率之间的关系.

命题 4.3 $P(E \cup F) = P(E) + P(F) - P(EF)$

证明：注意到 $E \cup F$ 可以表示为两个不相容事件 E 和 E^cF 的并, 根据公理 3 可知

$$P(E \cup F) = P(E \cup E^cF) = P(E) + P(E^cF)$$

另外, 由 $F = EF \cup E^cF$, 再利用公理 3 也可以得到

$$P(F) = P(EF) + P(E^cF)$$

或等价地,

$$P(E^cF) = P(F) - P(EF)$$

将它代入前面关于 $P(E \cup F)$ 的表达式, 命题得证.

命题 4.3 也可以利用韦恩图来证明, 如图 2.4.

将 $E \cup F$ 分成 3 个互不相容的部分, 如图 2.5 所示. 第 I 部分表示的是所有属于 E 但不属于 F 的点 (也即 EF^c), 第 II 部分表示的是所有既属于 E 也属于 F 的点 (也即 EF), 第 III 部分表示所有属于 F 但不属于 E 的点 (也即 E^cF).

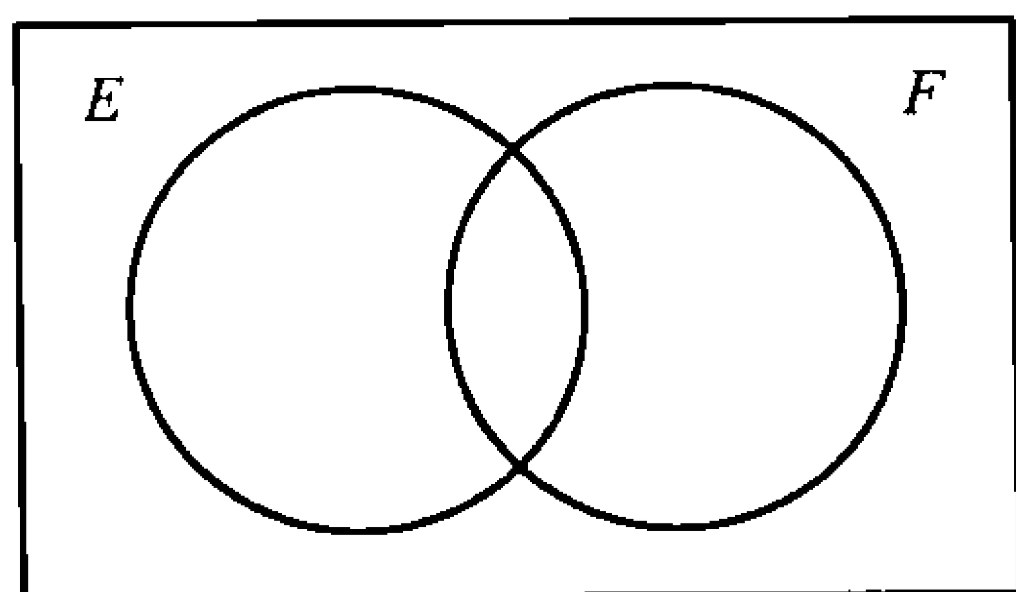


图2.4 韦恩图

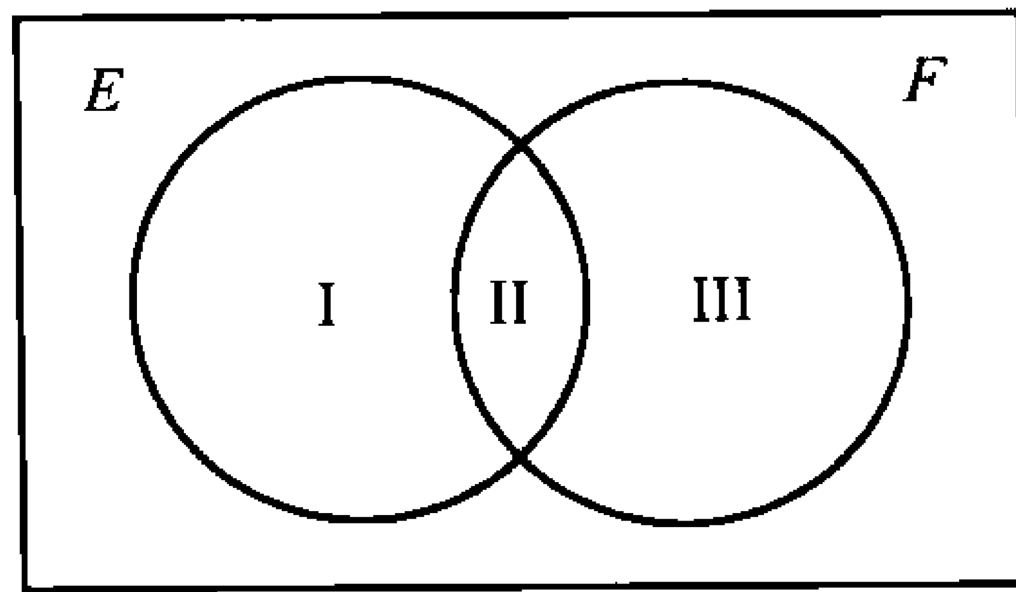


图2.5 韦恩图的3个部分

从图 2.5 上可以看出,

$$E \cup F = \text{I} \cup \text{II} \cup \text{III} \quad E = \text{I} \cup \text{II} \quad F = \text{II} \cup \text{III}$$

由于 I, II, III 是互不相容的, 结合公理 3 可以得到:

$$P(E \cup F) = P(\text{I}) + P(\text{II}) + P(\text{III}) \quad P(E) = P(\text{I}) + P(\text{II}) \quad P(F) = P(\text{II}) + P(\text{III})$$

由以上就可以得出 $P(E \cup F) = P(E) + P(F) - P(\text{II})$, 又因为 II=$EF$, 命题 4.3 得以证明. □

例 4a 某人度假时随身带了两本书. 他喜欢第一本书的概率为 0.5, 喜欢第二本书的概率为 0.4, 两本书都喜欢的概率为 0.3, 问两本书都不喜欢的概率是多大?

解：令 B_i 表示他喜欢第 i 本书 $(i = 1, 2)$, 那么他至少喜欢一本书的概率为

$$P(B_1 \cup B_2) = P(B_1) + P(B_2) - P(B_1B_2) = 0.5 + 0.4 - 0.3 = 0.6$$

因为两本书都不喜欢的对立事件是至少喜欢一本书, 所以

$$P(B_1^c B_2^c) = P\big((B_1 \cup B_2)^c\big) = 1 - P(B_1 \cup B_2) = 0.4 \qquad \blacksquare$$

以下公式计算三个事件 E, F, G 之中至少有一个发生的概率:

$$P(E \cup F \cup G) = P[(E \cup F) \cup G]$$

由命题 4.3 可知上式等于

$$P(E \cup F) + P(G) - P[(E \cup F)G]$$

由分配律可知 $(E \cup F)G = EG \cup FG$, 因此由上面式子可得

$$\begin{aligned}
P(E \cup F \cup G) &= P(E) + P(F) - P(EF) + P(G) - P(EG \cup FG) \\
&= P(E) + P(F) - P(EF) + P(G) - P(EG) - P(FG) + P(EGFG) \\
&= P(E) + P(F) + P(G) - P(EF) - P(EG) - P(FG) + P(EFG)
\end{aligned}$$

以下的命题, 也称为**容斥恒等式**(inclusion-exclusion identity), 可由归纳法推导得到.

命题 4.4

$$\begin{aligned}
P(E_1 \cup E_2 \cup \cdots \cup E_n) = &\sum_{i=1}^{n} P(E_i) - \sum_{i_1 < i_2} P(E_{i_1} E_{i_2}) + \cdots \\
&+ (-1)^{r+1} \sum_{i_1 < i_2 < \cdots < i_r} P(E_{i_1} E_{i_2} \cdots E_{i_r}) \\
&+ \cdots + (-1)^{n+1} P(E_1 E_2 \cdots E_n)
\end{aligned}$$

其中, $\sum_{i_1 < i_2 < \cdots < i_r} P(E_{i_1} E_{i_2} \cdots E_{i_r})$ 表示对集合 $\{1, 2, \cdots, n\}$ 的所有大小为 r 的子集所对应的值求和, 和项一共包含 $\binom{n}{r}$ 项.

也就是说, n 个事件并的概率, 等于这些事件的概率之和, 减去两个事件同时发生的概率之和, 再加上三个事件同时发生的概率之和……

注释 1. 作为命题 4.4 的直观解释, 首先注意到如果样本空间里的某个结果不属于任意的 E_i, 那么等式两边都不应该有它的概率. 另一方面, 假设某个结果正好包含在 m 个 E_i 里面 (其中 $m > 0$), 那么既然它属于 $\bigcup_i E_i$, 这个结果的概率在 $P(\bigcup_i E_i)$ 中只计算一次. 而且, 因为这个结果也被包含在形如 $E_{i_1}, E_{i_2}, \cdots, E_{i_k}$ 这样的 $\binom{m}{k}$ 个子集中, $k = 1, \cdots, m$, 在命题 4.4 等式的右边, 这个结果的概率被计算了

$$\binom{m}{1} - \binom{m}{2} + \binom{m}{3} - \cdots \pm \binom{m}{m}$$

次. 因此, 对于 $m>0$, 我们必须证明

$$1=\binom{m}{1}-\binom{m}{2}+\binom{m}{3}-\cdots\pm\binom{m}{m}$$

然而, 因为 $1=\binom{m}{0}$, 上式等价于

$$\sum_{i=0}^{m}\binom{m}{i}(-1)^i=0$$

而这是二项式定理的结果, 因为

$$0=(-1+1)^m=\sum_{i=0}^{m}\binom{m}{i}(-1)^i(1)^{m-i}$$

2. 以下式子是容斥恒等式更简明的写法:

$$P\Big(\bigcup_{i=1}^{n}E_i\Big)=\sum_{r=1}^{n}(-1)^{r+1}\sum_{i_1<\cdots<i_r}P(E_{i_1}\cdots E_{i_r})$$

3. 在容斥恒等式中, 右边如果只取前一项, 那么得到事件并的概率的一个上界; 如果取前两项, 得到事件并的概率的一个下界; 取前 3 项, 得到一个上界; 取前 4 项, 得到一个下界, 以此类推. 也就是说, 对于事件 $E_1,\cdots,E_n$, 有

$$P\Big(\bigcup_{i=1}^{n}E_i\Big)\leqslant\sum_{i=1}^{n}P(E_i) \tag{4.1}$$

$$P\Big(\bigcup_{i=1}^{n}E_i\Big)\geqslant\sum_{i=1}^{n}P(E_i)-\sum_{j<i}P(E_iE_j) \tag{4.2}$$

$$P\Big(\bigcup_{i=1}^{n}E_i\Big)\leqslant\sum_{i=1}^{n}P(E_i)-\sum_{j<i}P(E_iE_j)+\sum_{k<j<i}P(E_iE_jE_k) \tag{4.3}$$

等等. 为了证明这些不等式, 先注意到

$$\bigcup_{i=1}^{n}E_i=E_1\cup E_1^{c}E_2\cup E_1^{c}E_2^{c}E_3\cup\cdots\cup E_1^{c}\cdots E_{n-1}^{c}E_n$$

也就是说, E_i 里至少有一个发生, 相当于 E_1 发生, 或者 E_1 不发生, 但是 E_2 发生, 或者 E_1、E_2 都不发生, 但 E_3 发生, 等等. 因为上式的右边是一系列互不相容事件的并, 我们有

$$\begin{aligned}P\left(\bigcup_{i=1}^{n}E_i\right)&=P(E_1)+P(E_1^{c}E_2)+P(E_1^{c}E_2^{c}E_3)+\cdots+P(E_1^{c}\cdots E_{n-1}^{c}E_n)\\&=P(E_1)+\sum_{i=2}^{n}P(E_1^{c}\cdots E_{i-1}^{c}E_i)\end{aligned} \tag{4.4}$$

令 $B_i = E_1^c \cdots E_{i-1}^c = \left(\bigcup_{j<i} E_j\right)^c$ 表示前 $i-1$ 个事件都不发生, 利用恒等式

$$P(E_i) = P(B_i E_i) + P(B_i^c E_i)$$

可证明

$$P(E_i) = P(E_1^c \cdots E_{i-1}^c E_i) + P\Big(E_i \bigcup_{j<i} E_j\Big)$$

也

$$P(E_1^c \cdots E_{i-1}^c E_i) = P(E_i) - P\Big(\bigcup_{j<i} E_i E_j\Big)$$

将此代入 (4.4) 式即可得到

$$P\Big(\bigcup_{i=1}^{n} E_i\Big) = \sum_i P(E_i) - \sum_i P\Big(\bigcup_{j<i} E_i E_j\Big) \tag{4.5}$$

因为概率总是非负的, 所以, 由 (4.5) 式便可直接得到不等式 (4.1). 然后, 给定 i, 利用不等式 (4.1) 可得

$$P\Big(\bigcup_{j<i} E_i E_j\Big) \leqslant \sum_{j<i} P(E_i E_j)$$

此式结合 (4.5) 式, 又可得到 (4.2) 式. 现在给定 i, 将不等式 (4.2) 应用到 $P(\bigcup_{j<i} E_i E_j)$, 可得到

$$\begin{aligned} P\Big(\bigcup_{j<i} E_i E_j\Big) &\geqslant \sum_{j<i} P(E_i E_j) - \sum_{k<j<i} P(E_i E_j E_i E_k) \\ &= \sum_{j<i} P(E_i E_j) - \sum_{k<j<i} P(E_i E_j E_k) \end{aligned}$$

由上式, 结合 (4.5) 式, 便可得到 (4.3) 式. 其他的不等式都可以通过这种方法得到.

2.5 等可能结果的样本空间

对很多试验来说, 一个很自然的假设是, 样本空间里的所有结果发生的可能性是一样的. 也即, 假定样本空间 S 是个有限集, 不妨设为 $S = \{1, 2, \cdots, N\}$, 这时就经常会自然地假设

$$P(\{1\}) = P(\{2\}) = \cdots = P(\{N\})$$

结合公理 2 和公理 3, 上式意味着 (为什么?)

$$P(\{i\}) = \frac{1}{N} \quad i = 1, 2, \cdots, N$$

再利用公理 3 就可以得到对任何事件 E

$$P(E)=\frac{E\text{中的结果数}}{S\text{中的结果数}}$$

也就是说, 如果假定一次试验里的所有结果都是等可能发生的, 那么任何事件 E 发生的概率等于 E 中所含有的结果数占所有样本空间里的结果数的比例.

例 5a 掷两枚骰子, 朝上那一面数字之和为 7 的概率是多少?

解: 仍假设所有的 36 种可能结果都是等可能发生的, 这样, 就有 6 种可能结果满足数字之和等于 7, 即 (1,6),(2,5),(3,4),(4,3),(5,2),(6,1). 所以, 两枚骰子点数之和为 7 的概率应该是 $6/36=1/6$.

例 5b 一个碗里面一共有 6 个白球, 5 个黑球, 随机地从里面取出 3 个球, 问恰好有一个白球两个黑球的概率是多少?

解: 首先考虑取球是有顺序的, 样本空间一共包含 $11\times10\times9=990$ 种结果. 现在考虑事件"取出一个白球, 两个黑球"所包含的可能结果: 第一个球是白的, 后两个球是黑的一共有 $6\times5\times4=120$ 种; 第一个球是黑的, 第二个球是白的, 第三个球又是黑的一共有 $5\times6\times4=120$ 种; 前两个球是黑的, 第三个球是白的一共有 $5\times4\times6=120$ 种. "随机取"意味着样本空间的结果都是等可能发生的, 所以, 所求概率为 $\dfrac{120+120+120}{990}=\dfrac{4}{11}$.

这个问题也可以认为取球是没有顺序的, 从这个角度看, 样本空间一共存在 $\binom{11}{3}=165$ 种结果. 当然, 这 165 种结果也是等可能的. 与事件"一个白球, 两个黑球"相关的结果有 $\binom{6}{1}\binom{5}{2}$ 种, 因此, 抽出一个白球和两个黑球的概率为 $\binom{6}{1}\binom{5}{2}\Big/\binom{11}{3}=\dfrac{4}{11}$. 这个结果同前面答案是一致的.

当涉及从 n 个对象中随机地选取 k 个对象的时候, 我们可以认为对象是从 n 个对象的集合中一个一个地有序地选出来的, 也可以认为从 n 个对象中一次取出 k 个对象. 当一个一个地取出来的时候, 我们假定每次选择对象的时候, 在待选的集合中的每一个对象有相同的概率被选中. 在一次选择 k 个对象的时候, 我们假定 $\binom{n}{k}$ 个可能的子集有相同的机会被选中. 例如, 假定一个集体由 10 对夫妇组成, 现在从中随机地选择 5 人. 我们希望计算 $P(N)$, 其中事件 N 表示这 5 人中没有一对夫妇. 首先, 这 20 人中一共有 $\binom{20}{5}$ 个不同的人员组合, 每一个可能的人员组合都有相同的机会被选中. 而 N 这个事件中的结果可以看成 6 阶段的试验的结果, 第一阶段从 10 对夫妇中选定 5 对夫妇, 后面的 5 个阶段是顺次地从每一对夫妇中选出一人. 这样的试验共有 $\binom{10}{5}2^5$ 种不同的结果. 这样, N 的概率为

$$P(N)=\frac{\binom{10}{5}2^5}{\binom{20}{5}}.$$

我们也可以从顺序地选择这样的观点来计算 $P(N)$. 从 20 个人中顺序地选择 5 个人, 一共有 20·19·18·17·16 种等概的试验结果. 但是选择 5 个没有夫妇关系的 5 人组, 只有 20·18·16·14·12 个试验结果. 这样

$$P(N)=\frac{20\cdot18\cdot16\cdot14\cdot12}{20\cdot19\cdot18\cdot17\cdot16}.$$

经过简单的计算, 这两个结论是相同的, 至于具体的计算就留作练习吧! ■

例 5c 一个委员会由 5 人组成, 需要从 6 个男人和 9 个女人中随机选取, 问委员会由 3 个男人和 2 个女人组成的概率有多大?

解: 假定随机选取则意味着所有 $\binom{15}{5}$ 种组合的选择是等可能的, 而与事件 "3 男 2 女" 相关的结果有 $\binom{6}{3}\binom{9}{2}$ 种, 因此所讨论事件的概率为: $\binom{6}{3}\binom{9}{2}\Big/\binom{15}{5}$ $=\dfrac{240}{1001}$. ■

例 5d 一个坛子里共 n 个球, 其中一个做了标记. 如果依次从中随机抽取 k 个球, 问做了标记的球被取出来的概率有多大?

解: 从 n 个球中选取 k 个球, 一共有 $\binom{n}{k}$ 种选取方法, 每一种选取方法都是等可能的. 与事件"选中带标记的球"相关的选法共有 $\binom{1}{1}\binom{n-1}{k-1}$ 种, 因此

$$P\{\text{标记球被取出}\}=\frac{\binom{1}{1}\binom{n-1}{k-1}}{\binom{n}{k}}=\frac{k}{n}$$

也可以这样求解: 设 k 个球是顺序地被取出, 记 A_i 表示标记球在第 i 次被取出 $(i=1,2,\cdots,k)$. 既然所有球在第 i 次被抽取的概率是一样的, 可知 $P(A_i)=1/n$. 而这些事件是彼此互不相容的, 因此,

$$P\{\text{标记球被取出}\}=P\Big(\bigcup_{i=1}^{k}A_i\Big)=\sum_{i=1}^{k}P(A_i)=\frac{k}{n}$$

另外, $P(A_i)=1/n$ 可以这样推导: 考虑到抽球的过程是有顺序的, 一共有 $n(n-1)\cdots(n-k+1)=n!/(n-k)!$ 种等可能试验结果, 其中有 $(n-1)(n-2)\cdots(n-i+1)(1)(n-i)\cdots(n-k+1)=(n-1)!/(n-k)!$ 种试验结果表示标记

球在第 i 次抽取时被取出, 因此

$$P(A_i) = \frac{(n-1)!}{n!} = \frac{1}{n}$$ ■

例 5e 有 $n+m$ 个球, 其中 n 个红的, m 个蓝的, 将它们随机排成一排, 也即所有 $(n+m)!$ 种排列都是等可能的. 如果只顺序地记录下所排列的球的颜色, 证明将球的颜色序列作为试验结果的时候, 这些试验结果仍然是等概的.

解: 我们将 $(n+m)$ 个球的次序排列称为一组球的排列, 将 $n+m$ 个球的颜色次序排列称为一组球的颜色次序排列. 球的排列共有 $(n+m)!$ 种, 将红球之间作任何一个位置的置换, 将蓝球之间作任一位置置换, 置换的结果并不影响球的颜色次序排列. 从而, 一组球的颜色次序排列, 对应于 $n!m!$ 个球的排列. 这说明球的次序排列也是等可能的. 并且每一种颜色次序出现的概率为 $n!m!/(n+m)!$.

例如, 假设有 2 个红球, 记为 $\mathrm{r}_1, \mathrm{r}_2$; 两个蓝球, 记为 $\mathrm{b}_1, \mathrm{b}_2$; 这样, 一共有 4! 种球的排列, 每一个颜色次序的排列, 对应于 2!2! 个球的排列. 如下面四个球的排列对应于相同的颜色次序排列:

$$\mathrm{r}_1, \mathrm{b}_1, \mathrm{r}_2, \mathrm{b}_2 \qquad \mathrm{r}_1, \mathrm{b}_2, \mathrm{r}_2, \mathrm{b}_1 \qquad \mathrm{r}_2, \mathrm{b}_1, \mathrm{r}_1, \mathrm{b}_2 \qquad \mathrm{r}_2, \mathrm{b}_2, \mathrm{r}_1, \mathrm{b}_1$$

因此, 每一个颜色次序排列出现的概率为 $4/24 = 1/6$. ■

例 5f 在扑克牌游戏中, 一手牌有 5 张, 如果这 5 张牌是连续的, 但又不是同一花色, 那么称为顺子, 比如, "黑桃 5, 黑桃 6, 黑桃 7, 黑桃 8, 红桃 9"就是一副顺子. 试求一手牌是顺子的概率是多大?

解: 假设所有 $\binom{52}{5}$ 种组合都是等可能的. 先看看由"A,2,3,4,5"这 5 张牌 (花色不同) 能组成多少个顺子, 首先, "A"有 4 种可能, 同样其他 4 张牌也分别有 4 种可能, 所以, 一共有 4^5 种可能, 但是, 其中有 4 种可能是 5 张牌是同花色 (这种情况称为同花顺), 所以一共是 4^5-4 种顺子. 类似地, "10, J, Q, K, A"这种顺子也有 4^5-4 种, 因此一共有 $10\times(4^5-4)$ 种顺子. 这样所求概率为 $10\times(4^5-4)\Big/\binom{52}{5} \approx 0.0039$. ■

例 5g 一手牌有 5 张, 如果其中 3 张点数一样, 另两张点数也一样的话, 称为"福尔豪斯"(full house, 也就是说"福尔豪斯"由 3 张点数一样的牌加上一对组成). 试问一手牌恰好是"福尔豪斯"的概率是多大?

解: 同样也设所有 $\binom{52}{5}$ 种组合都是等可能的. 注意到像"2 张 10, 3 张 J"这样的"福尔豪斯"一共有 $\binom{4}{2}\binom{4}{3}$ 种组合, 又因为一对的点数有 13 种选择, 在选定一对的点数后, 剩下 12 种可能的点数用于选择 3 张一组的牌. 所以所求概率为

$$13\times 12\times\binom{4}{2}\binom{4}{3}\Big/\binom{52}{5} \approx 0.0014$$ ■

例 5h 桥牌比赛中, 52 张牌被分给 4 位选手, 求下列事件概率:

(a) 其中有一人拿到了 13 张黑桃;

(b) 每人都拿到了 1 张“A”.

解: (a) 记 E_i 为第 i 个人拿到这 13 张黑桃的事件, 显然

$$P(E_i) = \frac{1}{\binom{52}{13}}, \quad i = 1, 2, 3, 4$$

由于 E_i $(i = 1, 2, 3, 4)$ 是互不相容的事件, 其中某一人拿到这 13 张黑桃的概率为

$$P\left(\bigcup_{i=1}^{4} E_i\right) = \sum_{i=1}^{4} P(E_i) = 4\Big/\binom{52}{13} \approx 6.3 \times 10^{-12}$$

(b) 现来求每个选手恰好拿一张“A”的概率. 先把“A”放一边, 剩下 48 张牌分给 4 个人的可能分派方法数为 $\binom{48}{12, 12, 12, 12}$. 接下来, 将 4 张 A 分给 4 个选手的可能分派方法数为 4!, 因此, 每人得到 1 张 A 的所有可能分派方法数为 $4! \times \binom{48}{12, 12, 12, 12}$. 从而所求概率为

$$4! \times \binom{48}{12, 12, 12, 12}\Big/\binom{52}{13, 13, 13, 13} \approx 0.105\ 5$$ ■

有些事件的概率是出乎想象的, 下面两个例子就是如此.

例 5i 房间里有 n 个人, 没有两人在同一天生日的概率是多大? n 多大时, 才能保证此概率小于 1/2?

解: 每个人的生日都有 365 种可能, 所以 n 个人一共是 365^n 种可能 (此处忽略有人生日是 2 月 29 日的可能性). 假定每种可能性都是一样的. 可知所求事件的概率为 $(365)(364)(363)\cdots(365-n+1)/365^n$. 令人惊异的是, 一旦 $n \geqslant 23$, 这个概率就比 1/2 要小, 也就是说, 房间里人数如果超过 23 的话, 那么至少有两人为同一天生日的概率就大于 1/2. 很多人一开始对这个结果很吃惊, 因为 23 相对于一年 365 天来说太小了, 然而, 对每两个人来说, 生日相同的概率为 $\frac{365}{(365)^2} = \frac{1}{365}$, 23 个人, 一共可以组成 $\binom{23}{2} = 253$ 对, 这样看来, 上述结果就不会太令人吃惊了.

当房间里人数达到 50 时, 至少两人同一天生日的概率大概为 97%, 如果人数达到 100, 那么两人同一天生日的优势 (优势的定义见 3.3 节) 为 $3 \times 10^6 : 1$, 也就是说这个概率比 $\frac{3 \times 10^6}{3 \times 10^6 + 1}$ 还要大. ■

例 5j 52 张牌扣在桌上一张一张翻开, 一直到出现一张“A”为止. 接下来再翻一张牌, 问出现黑桃“A”和出现梅花 2 的概率哪个大?

解: 为了计算第一张“A”之后出现黑桃“A”的概率, 先要计算在所有的 (52)! 种可能的发牌次序中, 有多少种是第一次出现“A”后紧接着就出现黑桃“A”. 先将 51 张牌 (去掉黑桃“A”) 任意排列, 然后将黑桃“A”找个位置插进去, 然而, 只有一个位置即第一次出现“A”以后接下来的位置才是满足要求的位置. 比如, 假设其他 51 张牌的顺序为:

梅花 4, 红桃 6, 方块“J”, 黑桃 5, 梅花“A”, 方块 7, ……, 红桃“K”

接下来, 只有一种插入方法满足条件, 即:

梅花 4, 红桃 6, 方块“J”, 黑桃 5, 梅花“A”, 黑桃“A”, 方块 7, ……, 红桃“K”

因此, 第一张 A 出现后, 紧接着是黑桃“A”的次序一共有 (51)! 种, 而所求概率为

$$P(\text{第一张 A 后是黑桃“A”}) = \frac{(51)!}{(52)!} = \frac{1}{52}$$

用同样的方法, 可以得知, 第一张“A”后出现梅花 2(或任何其他牌) 的概率也为 1/52. 也就是说, 52 张牌中的任意一张牌 (包括任意花色的“A”) 出现在第一张“A”后面的概率都是 1/52.

这个结果会让很多人吃惊! 事实上, 一般的反应都是认为第一个“A”出现后, 接着出现梅花 2 的概率要大于出现黑桃“A”的概率. 因为第一张“A”就有可能是黑桃“A”本身. 这就减少了黑桃“A”紧接着第一张“A”的可能性. 但是他忽略了第一张“A”前面可以出现梅花 2 这一事实. 这又减少了梅花 2 紧跟在第一张“A”后的可能性. 然而, 因为有 1/4 的机会黑桃“A”是第一张出现的“A”(因为 4 张“A”出现在第一位是等可能的), 而且, 仅仅有 1/5 的机会是梅花 2 出现在第一张“A”之前 (因为梅花 2 和四张“A”中的任一张排在最前面的可能性是相同的), 这点又好像说明了梅花 2 紧跟在第一张“A”后的可能要大一些. 然而, 事实并非如此, 更深入的分析可以说明两者的可能性是一样的. ■

例 5k 一个橄榄球队有 20 名进攻球员和 20 名防守球员, 现在要给他们安排宿舍, 每两人一个宿舍. 如果随机分派, 没有一个宿舍既有进攻队员也有防守队员 (简称为“进攻防守对”) 的概率是多大? 正好有 $2i$ 对“进攻防守对”的概率是多大 $(i = 1, 2, \cdots, 10)$?

解: 将 40 人分成有序的 20 对一共有 $\binom{40}{2,2,\cdots,2} = \frac{(40)!}{(2!)^{20}}$ 种可能. 也就是说, 一共有 $(40)!/2^{20}$ 种方法将这些选手分成第一对、第二对, 等等. 因此, 一共有 $(40)!/[2^{20}(20)!]$ 种方法分成不考虑顺序的 20 对. 而且, 要想不出现“进攻防守对”, 只有将进攻队员之间配对, 防守队员之间配对, 一共有 $[(20)!/2^{10}(10)!]^2$ 种方法. 因此, 没有“进攻防守对”的概率 (记为 P_0) 如下:

$$P_0=\frac{\left(\dfrac{(20)\,!}{2^{10}(10)\,!}\right)^2}{\dfrac{(40)\,!}{2^{20}(20)\,!}}=\frac{[(20)\,!]^3}{[(10)\,!]^2(40)\,!}$$

现在计算 P_{2i}, 也即正好有 $2i$ 对“进攻防守对”的概率. 首先, 一共有 $\binom{20}{2i}^2$ 种方法选取 $2i$ 个进攻队员和 $2i$ 个防守队员以便组成“进攻防守对”, 这 $4i$ 个人能够组成 $(2i)\,!$ 种可能“进攻防守对”(因为第一个进攻队员可以和 $2i$ 个防守队员配对, 第二个进攻队员可以和 $2i-1$ 个防守队员配对, 依此类推). 剩下的 $20-2i$ 个进攻队员 (防守队员) 只能内部配对. 一共有 $\binom{20}{2i}^2(2i)\,!\left[\dfrac{(20-2i)\,!}{2^{10-i}(10-i)\,!}\right]^2$ 种可能方式配成“进攻防守对”. 因此

$$P_{2i}=\frac{\binom{20}{2i}^2(2i)\,!\left[\dfrac{(20-2i)\,!}{2^{10-i}(10-i)\,!}\right]^2}{\dfrac{(40)\,!}{2^{20}(20)\,!}}\qquad i=0,1,\cdots,10$$

这样, 根据上述公式就可以算出 $P_{2i}, i=0,1,\cdots,10$. 另外, 利用 Stirling 公式 ($n\,!\approx n^{n+1/2}\mathrm{e}^{-n}\sqrt{2\pi}$) 还可以算出其估值. 比如

$$P_0\approx 1.340\,3\times10^{-6}\qquad P_{10}\approx 0.345\,861\qquad P_{20}\approx 7.606\,8\times10^{-6}$$ ■

接下来三个例子是命题 4.4 的运用. 在例 5l 中引入概率来迅速解决计数问题.

例 5l 一个俱乐部里有 36 人会打网球, 28 人会打软式网球, 18 人会打羽毛球; 22 人会打网球和软式网球, 12 人会打网球和羽毛球, 9 人会打软式网球和羽毛球; 4 人三种球都会打. 试问: 至少会打一种球的有多少人?

解: 记 N 为俱乐部总人数. 设从俱乐部中随机地抽取一人, 又假设 C 为它的任一子集, 那么抽到一人刚好在 C 中的概率为

$$P(C)=\frac{C\text{中人数}}{N}$$

设 T,S,B 分别表示会打网球、软式网球和羽毛球的人的集合, 那么利用上述公式以及命题 4.4 可知

$$\begin{aligned}&P(T\cup S\cup B)\\=&P(T)+P(S)+P(B)-P(TS)-P(TB)-P(SB)+P(TSB)\\=&\frac{36+28+18-22-12-9+4}{N}=\frac{43}{N}\end{aligned}$$

因此, 至少会打一种球的人数为 43 人. ■

接下来的例子不但有一个很让人吃惊的答案, 而且在理论上也很有意义.

例 5m (配对问题) 房间里有 N 人参加舞会. 如果所有人都将帽子扔到屋中

央混在一起, 然后每人再随机拿一个帽子. 求没有一个人拿到自己的帽子的概率.

解: 先计算至少有一人拿到自己的帽子的概率. 记 $E_i, i=1,2,\cdots,N$ 表示"第 i 人拿到了自己的帽子". 这样, 由命题 4.4, 至少有一人拿到了自己的帽子的概率为:

$$\begin{aligned}P\Big(\bigcup_{i=1}^{N}E_i\Big)=&\sum_{i=1}^{N}P(E_i)-\sum_{i_1<i_2}P(E_{i_1}E_{i_2})+\cdots\\&+(-1)^{n+1}\sum_{i_1<i_2<\cdots<i_n}P(E_{i_1}E_{i_2}\cdots E_{i_n})\\&+\cdots+(-1)^{N+1}P(E_1E_2\cdots E_N)\end{aligned}$$

如果把试验结果看成是一个 N 维向量, 其中第 i 个元素是第 i 个人拿到的帽子编号, 这样一共有 $N!$ 种可能的结果, 例如, 结果 $(1,2,3,\cdots,N)$ 表示每人拿到的都是自己的帽子. 进一步, $E_{i_1}E_{i_2}\cdots E_{i_n}$ 表示 $i_1,i_2,\cdots,i_n$ 这 n 个人拿到的是自己的帽子, 这种可能性会有 $(N-n)(N-n-1)\cdots3\cdot2\cdot1=(N-n)!$ 种, 因为剩下的 $N-n$ 个人, 第一个人有 $N-n$ 种选择方法, 第二人有 $N-n-1$ 种选择方法, 依此类推. 由假定知 N 个人的 $N!$ 种选择都是等可能的, 因此

$$P(E_{i_1}E_{i_2}\cdots E_{i_n})=\frac{(N-n)!}{N!}$$

又因 $\sum_{i_1<i_2<\cdots<i_n}P(E_{i_1}E_{i_2}\cdots E_{i_n})$ 一共含有 $\binom{N}{n}$ 项, 所以

$$\sum_{i_1<i_2<\cdots<i_n}P(E_{i_1}E_{i_2}\cdots E_{i_n})=\frac{N!(N-n)!}{(N-n)!n!N!}=\frac{1}{n!}$$

将上式代入 $P\big(\bigcup_{i=1}^{n}E_i\big)$ 的公式, 得

$$P\Big(\bigcup_{i=1}^{N}E_i\Big)=1-\frac{1}{2!}+\frac{1}{3!}-\cdots+(-1)^{N+1}\frac{1}{N!}$$

因此, 没有一人拿到自己的帽子的概率为:

$$1-1+\frac{1}{2!}-\frac{1}{3!}+\cdots+(-1)^{N}\frac{1}{N!}$$

当 N 足够大时, 上值约等于 $e^{-1}\approx0.367\,8$, 即当 N 很大时, 没有一人拿到自己的帽子的概率接近 0.37. (有多少人会误地认为, 当 $N\to\infty$ 时, 此概率值会趋近 1?) ■

以下的例子展示了命题 4.4 的又一个应用.

例 5n 10 对夫妇坐成一圈, 计算没有一对夫妻坐在一起的概率.

解: 记 E_i $(i=1,2,\cdots,10)$ 表示第 i 对夫妇坐在一起, 因此, 所求概率为 $1-P\Big(\bigcup_{i=1}^{10}E_i\Big)$.

由命题 4.4,

$$P\left(\bigcup_{i=1}^{10} E_i\right) = \sum_{i=1}^{10} P(E_i) - \cdots + (-1)^{n+1} \sum_{i_1<i_2<\cdots<i_n} P(E_{i_1}E_{i_2}\cdots E_{i_n}) + \cdots - P(E_1E_2\cdots E_{10})$$

为了计算 $P(E_{i_1}E_{i_2}\cdots E_{i_n})$, 先注意到 20 个人坐成一圈, 一共有 19! 种可能. (为什么?) 对于指定的 n 对夫妇, 在排位时, 为使这 n 对夫妇坐在一起, 先把这 n 对夫妇中的每一对夫妇看成一个整体, 这样, 在排位时一共有 $20-n$ 个对象, 在圆桌上一共有 $(20-n-1)!$ 种排位的方法, 当排位确定以后, 这 n 对夫妇之间又有排位问题, 是男左女右还是男右女左. 这样, 将这 n 对夫妇排在一起的排位方法一共有 $2^n(19-n)!$ 种, 我们得到

$$P(E_{i_1}E_{i_2}\cdots E_{i_n}) = \frac{2^n(19-n)!}{(19)!}$$

再由命题 4.4, 可以得到至少有一对夫妇是坐在一起的概率为

$$\binom{10}{1}2^1\frac{(18)!}{(19)!} - \binom{10}{2}2^2\frac{(17)!}{(19)!} + \binom{10}{3}2^3\frac{(16)!}{(19)!} - \cdots - \binom{10}{10}2^{10}\frac{9!}{(19)!} \approx 0.660\,5$$

所有的妻子都不坐在她丈夫旁边的概率大约为 0.339 5. ■

*例 5o (游程) 假设某个赛季过后, 田径队的成绩为 n 胜 m 败. 通过对这个输和赢的序列的研究, 可望得到关于田径队的潜力的进一步的信息. 一种办法是研究输和赢的游程的规律. 所谓赢的一个游程就是指赢的连续序列. 例如, 如果 $n=10, m=6$, 这个序列为 WWLLWWWLWLLLWWWW, 其中 W 为赢, L 为输. 那么这里有 4 个赢的游程, 第一个游程的大小为 2, 第二个游程的大小为 3, 第三个游程的大小为 1, 第四个游程的大小为 4.

现在假定这个球队有 n 次赢, m 次输. 假定所有的 $(n+m)!/(n!m!)$ 种次序是等可能的. 我们希望求出球队输赢的序列具有 r 个赢的游程的概率. 为此, 考虑满足条件 $x_1+\cdots+x_r=n$ 的正整数解 $x_1,\cdots,x_r$ 所组成的向量. 现在我们观察, 有多少个输赢序列满足如下条件: (i) 具有 r 个赢的游程, (ii) 第一个赢的游程的大小为 x_1, ……, 第 r 个赢的游程的大小为 x_r. 为此, 我们令 y_1 为第一个赢的游程以前输的次数, y_2 为第一个赢的游程与第二个赢的游程之间输的次数, ……, y_{r+1} 为最后一个赢的游程后面输的次数. 这些 y_i 满足

$$y_1+y_2+\cdots+y_{r+1}=m \qquad y_1\geqslant 0, y_{r+1}\geqslant 0, y_i>0, i=2,\cdots,r$$

这些 x_i, y_i 与相应的序列可以用下列图形形象地表示:

$$\underbrace{\text{LL}\cdots\text{L}}_{y_1}\underbrace{\text{WW}\cdots\text{W}}_{x_1}\underbrace{\text{L}\cdots\text{L}}_{y_2}\underbrace{\text{WW}\cdots\text{W}}_{x_2}\cdots\underbrace{\text{WW}}_{x_r}\underbrace{\text{L}\cdots\text{L}}_{y_{r+1}}$$

现在可以数一数, 对于固定的 $x_1,\cdots,x_r$, 相应输序列的个数为向量 $(y_1,\cdots,y_{r+1})$ 的个数, 其中 $y_1,\cdots,y_{r+1}$ 满足前面所提到的约束条件. 为了进一步计算输序列的个数, 令

$$\bar{y}_1=y_1+1,\quad \bar{y}_i=y_i,\quad i=2,\cdots,r,\quad \bar{y}_{r+1}=y_{r+1}+1$$

输序列的个数变成满足下列条件

$$\bar{y}_1+\bar{y}_2+\cdots+\bar{y}_{r+1}=m+2$$

的正整数向量 $(\bar{y}_1,\cdots,\bar{y}_{r+1})$ 的个数, 在第 1 章命题 6.1 中, 指出这个方程的正整数解的个数为 $\binom{m+1}{r}$. 这样, 具有 r 个赢的游程的输赢序列的个数为 $\binom{m+1}{r}$ 乘以 $x_1+\cdots+x_r=n$ 的正整数解的个数. 再一次利用第 1 章命题 6.1 的结论, 得到如下结论: 具有 r 个赢的游程的输赢序列的个数为 $\binom{m+1}{r}\binom{n-1}{r-1}$. 由于我们假定 $\binom{n+m}{n}$ 个输赢序列是等可能的, 故

$$P\{\text{赢的游程的个数为}r\}=\binom{m+1}{r}\binom{n-1}{r-1}\Big/\binom{m+n}{n}\qquad r\geqslant 1$$

例如, $n=8, m=6$ 则具有 7 个赢的游程的概率为 $\binom{7}{7}\binom{7}{6}\Big/\binom{14}{8}=1/429$. 此处假设所有的 $\binom{14}{8}$ 个输赢序列是等可能的. 现在假定这个队的输赢结果是 WLWLWLWLWWLWLW. $r=7$ 发生的概率很小, 此时, 我们可能会认为输赢的概率会随着时间变化, 往往输球以后赢球的概率较大, 而赢球以后输球的概率大. 排球比赛中容易出现这种情况, 因为赢球以后, 保持了发球权, 这样会给对方进攻的机会. 另一方面, 若输赢的结果是 WWWWWWWWLLLLLL, 此时, $P\{\text{赢的游程的个数为 }1\}=\binom{7}{1}\binom{7}{0}\Big/\binom{14}{8}=1/429$. 这种情况下, 我们就要怀疑球队的状况在下滑. ■

*2.6 概率：连续集函数

一列事件 $\{E_n, n\geqslant 1\}$ 称为递增列, 如果

$$E_1\subset E_2\subset\cdots\subset E_n\subset E_{n+1}\subset\cdots$$

反之, 称为递减列, 如果

$$E_1\supset E_2\supset\cdots\supset E_n\supset E_{n+1}\supset\cdots$$

如果 $\{E_n, n\geqslant 1\}$ 是递增事件列, 定义一个新的事件, 记为 $\lim\limits_{n\to\infty}E_n$, 如下:

$$\lim_{n\to\infty}E_n=\bigcup_{i=1}^{\infty}E_i$$

类似地, 如果 $\{E_n, n \geqslant 1\}$ 是递减事件列, 那么定义 $\lim\limits_{n\to\infty} E_n$, 如下:

$$\lim_{n\to\infty} E_n = \bigcap_{i=1}^{\infty} E_i$$

现在来证明命题 6.1.

命题 6.1 如果 $\{E_n, n \geqslant 1\}$ 是递增或者递减事件列, 那么

$$\lim_{n\to\infty} P(E_n) = P(\lim_{n\to\infty} E_n)$$

证明: 首先假设 $\{E_n, n \geqslant 1\}$ 是递增事件列, 并且定义事件 $F_n(n \geqslant 1)$ 如下:

$$F_1 = E_1 \quad F_n = E_n\left(\bigcup_{i=1}^{n-1} E_i\right)^{\mathrm{c}} = E_n E_{n-1}^{\mathrm{c}} \quad n > 1$$

此处用到了 $\bigcup_{i=1}^{n-1} E_i = E_{n-1}$, 也就是说 F_n 是由那些属于 E_n 但是不属于 $E_i(i < n)$ 的元素组成. 显然, F_n 是一列互不相容事件. 且满足

$$\bigcup_{i=1}^{\infty} F_i = \bigcup_{i=1}^{\infty} E_i \quad 以及 \quad \bigcup_{i=1}^{n} F_i = \bigcup_{i=1}^{n} E_i \quad 对任意的 n \geqslant 1$$

因此

$$\begin{aligned} P\left(\bigcup_{i=1}^{\infty} E_i\right) &= P\left(\bigcup_{i=1}^{\infty} F_i\right) = \sum_{i=1}^{\infty} P(F_i) \quad (利用公理 3) \\ &= \lim_{n\to\infty} \sum_{i=1}^{n} P(F_i) = \lim_{n\to\infty} P\left(\bigcup_{i=1}^{n} F_i\right) = \lim_{n\to\infty} P\left(\bigcup_{i=1}^{n} E_i\right) = \lim_{n\to\infty} P(E_n) \end{aligned}$$

这样就证明了 $\{E_n, n \geqslant 1\}$ 递增事件列时结论是成立的.

如果 $\{E_n, n \geqslant 1\}$ 是递减的, 那么 $\{E_n^{\mathrm{c}}, n \geqslant 1\}$ 是递增的, 因此, 根据上面结论可得到

$$P\left(\bigcup_{i=1}^{\infty} E_i^{\mathrm{c}}\right) = \lim_{n\to\infty} P(E_n^{\mathrm{c}})$$

由 $\bigcup_{i=1}^{\infty} E_i^{\mathrm{c}} = \left(\bigcap_{i=1}^{\infty} E_i\right)^{\mathrm{c}}$, 可知

$$P\left(\left(\bigcap_{i=1}^{\infty} E_i\right)^{\mathrm{c}}\right) = \lim_{n\to\infty} P(E_n^{\mathrm{c}})$$

等价地

$$1 - P\left(\bigcap_{i=1}^{\infty} E_i\right) = \lim_{n\to\infty} [1 - P(E_n)] = 1 - \lim_{n\to\infty} P(E_n)$$

即

$$P\left(\bigcap_{i=1}^{\infty} E_i\right) = \lim_{n\to\infty} P(E_n)$$

这样就证明了当 E_n 为递减事件列的时候命题的结论也成立. □

例 6a(概率与悖论) 设想有个无限大的坛子, 以及无限个编了号码 $1,2,3\cdots$ 的球, 进行以下的试验:

在 12 点差 1 分的时候, 将 1 到 10 号球放进去, 并把 10 号球拿出来 (假设放球和拿球的时间忽略不计);

在 12 点差 1/2 分的时候, 将 11 到 20 号球放进去, 并把 20 号球拿出来;

在 12 点差 1/4 分的时候, 将 21 号到 30 号球放进去, 并把 30 号球拿出来;

在 12 点差 1/8 分的时候, ……

等等. 问题: 在 12 点的时候, 坛子里有多少球?

问题的答案很明显: 12 点钟的时候坛子里有无限个球. 因为只要不是号码为 $10n, n \geqslant 1$ 的球, 都将在 12 点前放进去, 并且不会被取出来. 因此, 如果试验是这样进行的话, 问题已得到了解决.

现在换个试验:

在 12 点差 1 分的时候, 将 1 到 10 号球放进去, 并把 1 号球拿出来;

在 12 点差 1/2 分的时候, 将 11 到 20 号球放进去, 并把 2 号球拿出来;

在 12 点差 1/4 分的时候, 将 21 号到 30 号球放进去, 并把 3 号球拿出来;

在 12 点差 1/8 分的时候, ……

等等. 新的试验在 12 点钟的时候坛子里应该有多少球?

非常奇怪, 答案是在 12 点的时候, 坛子里一个球也没有. 理由是, 任何号码的球在 12 点前都将从坛子里取出, 比如号码为 n 的球, 在差 $(1/2)^{n-1}$ 到 12 点的时候被取出, 因此, 对于任意号码的某个球, 在 12 点的时候都不可能在坛子里, 这样, 坛子就是空的.

从上述讨论可以看出, 取球的方式不一样会导致结果不一样: 前一种情况, 只有号码为 $10n, n \geqslant 1$ 的球会被取出来, 但在后一种情况下, 所有的球都将被取出来.

现在设想在取球的时候, 是从当前所有球中随机取出, 也即:

在 12 点差 1 分的时候, 将 1 到 10 号球放进去, 并随机取一个球出来;

……

等等. 这种情况下, 在 12 点时, 坛子里有多少球?

解: 将要证明, 在 12 点时坛子为空的概率为 1.

首先考虑 1 号球, 定义 E_n 表示在"进行 n 次取球后, 1 号球仍在坛子里"这一事件. 很显然:

$$P(E_n) = \frac{9\cdot 18\cdot 27\cdots(9n)}{10\cdot 19\cdot 28\cdots(9n+1)}$$

注意到在经历 n 次取球后, 如果 1 号球仍在坛子里, 那么第一次取球有 9 种可能, 第 2 次取球有 18 种可能 (第二次取球的时候坛子里有 19 个球, 其中有一个是 1 号球), 等等. 这样, 12 点钟时, 1 号球仍在坛子里这一事件可以写为: $\bigcap_{n=1}^{\infty} E_n$, $E_n(n \geqslant 1)$ 为递减事件列, 根据命题 6.1 可知:

$$\begin{aligned}&P\{12 \text{ 点时 } 1 \text{ 号球仍在坛子里}\}\\ &=P\left(\bigcap_{n=1}^{\infty} E_n\right) = \lim_{n\to\infty} P(E_n) = \prod_{n=1}^{\infty}\left(\frac{9n}{9n+1}\right)\end{aligned}$$

现在证明

$$\prod_{n=1}^{\infty}\left(\frac{9n}{9n+1}\right) = 0$$

因为

$$\prod_{n=1}^{\infty}\left(\frac{9n}{9n+1}\right) = \left[\prod_{n=1}^{\infty}\left(\frac{9n+1}{9n}\right)\right]^{-1}$$

即等价于证明:

$$\prod_{n=1}^{\infty}\left(1+\frac{1}{9n}\right) = \infty$$

因为, 对任意 $m \geqslant 1$, 都有

$$\begin{aligned}\prod_{n=1}^{\infty}\left(1+\frac{1}{9n}\right) &\geqslant \prod_{n=1}^{m}\left(1+\frac{1}{9n}\right)\\ &= \left(1+\frac{1}{9}\right)\left(1+\frac{1}{18}\right)\left(1+\frac{1}{27}\right)\cdots\left(1+\frac{1}{9m}\right)\\ &> \frac{1}{9}+\frac{1}{18}+\frac{1}{27}+\cdots+\frac{1}{9m} = \frac{1}{9}\sum_{i=1}^{m}\frac{1}{i}\end{aligned}$$

因此, 令 $m\to\infty$ 且利用 $\sum_{i=1}^{\infty} 1/i = \infty$ 可以得到

$$\prod_{n=1}^{\infty}\left(1+\frac{1}{9n}\right) = \infty$$

令 F_i 表示"i 号球在 12 点时仍在坛子里"这一事件. 前面已证明 $P(F_1)=0$, 类似地, 可以证明对任意 i, $P(F_i)=0$. (比如, 同样的推理可以证明对任意 $i=11,12,\cdots,20$ 有 $P(F_i)=\prod_{n=2}^{\infty}[9n/(9n+1)]=0$). 因此, 12 点时坛子非空的概率为 $P\left(\bigcup_{i=1}^{\infty} F_i\right)$, 利用布尔不等式 (见自检习题 14) 可得:

$$P\left(\bigcup_{i=1}^{\infty} F_i\right) \leqslant \sum_{i=1}^{\infty} P(F_i) = 0$$

因此, 在 12 点时, 坛子为空的概率为 1. ■

2.7　概率：确信程度的度量

本文曾叙述：一个事件的概率，是指在重复进行某个试验的情况下，对该事件发生频率的一种度量．然而，概率还有另外的用处．比如，我们经常听到这样的评论，“90% 的可能是莎士比亚真地写了《哈姆雷特》”，“奥斯瓦尔德独自暗杀肯尼迪总统的可能性为 80%”，这又作何解释？

最自然又简单的解释是，概率是人们对自己的说法的确信程度的一种度量，也就是说，前面的陈述者比较确信“奥斯瓦尔德是独立行动的”，而且更加确信“莎士比亚写了《哈姆雷特》”．概率作为个体确信程度的度量这种解释经常被称为主观概率 (Subjective).

假设“确信程度的度量”满足概率的所有公理是很合情合理的．比如，如果我们有 70% 的把握认为是莎士比亚写了《凯撒大帝》，而只有 10% 的把握认为作者是马洛，那么我们应该有 80% 的把握认为作者是莎士比亚或是马洛．因此，无论把概率解释为确信程度的度量，还是事件发生的频率，其数学属性是不改变的．

例 7a　假设有 7 匹马参加比赛，而您认为 1 号马和 2 号马各有 20% 的机会获胜，3 号马和 4 号马各有 15% 的机会，其余 3 匹各 10% 的机会．如果进行同等赌注的押赌，是赌“获胜者将是 1,2,3 号马之一”还是赌“获胜者将是 1,5,6,7 号马之一”更好？

解：基于对比赛结果的主观认识，赌第一种赢的概率是：$0.2+0.2+0.15=0.55$，而第二种是 $0.2+0.1+0.1+0.1=0.5$，因此，赌第一种更好．■

应当指出，主观概率也应符合概率论的公理．但实际情况并非如此．例如，我们向某人了解对天气的看法时，经常提这样的问题：

(a) 今天下雨的可能性是多少？

(b) 明天下雨的可能性是多少？

(c) 今明两天都下雨的可能性是多少？

(d) 今天或明天会下雨的可能性是多少？

这个人经过考虑，很可能会给出下面的答案：30%,40%,20%,60%．显然，这样的回答与概率论的公理是相矛盾的．我们当然希望经过指出这种错误以后，这个人会修正他的回答，因为个人的看法不是精确计算出来的，是带有误差的．(一个可能可以接受的修正是：30%, 40%,10%,60%.)

小　结

如果令 S 为表示某个试验的所有可能结果的集合，那么 S 称为该试验的样本空间．一个事件就是 S 的一个子集．如果 $A_i, i=1,\cdots,n$ 为一系列事件，那么称

$\bigcup_{i=1}^n A_i$ 为这些事件的并, 它表示至少包含在某一个 A_i 里的所有结果所构成的事件. 类似地, $\bigcap_{i=1}^n A_i$ 称为这些事件的交, 有时也记为 $A_1 \cdots A_n$, 表示包含在所有 A_i 里的所有结果所构成的事件.

对任一事件 A, 定义 A^c 为由那些不包含在 A 里的所有结果所构成的事件, 称为 A 的**对立事件**. 事件 S^c 不包含任何结果, 记为 $\varnothing$, 称为**空集**. 如果 $AB = \varnothing$, 那么称 A 和 B **互不相容**.

设对于样本空间里的任一事件 A, 对应于一个数 $P(A)$, 若集合函数 $P(A)$ 满足以下条件, 则称 $P(A)$ 为 A 的概率:

(i) $0 \leqslant P(A) \leqslant 1$

(ii) $P(S) = 1$

(iii) 对于任意互不相容事件 $A_i, i \geqslant 1$, 有

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

$P(A)$ 表示试验结果包含在事件 A 里的概率.

容易证明:

$$P(A^c) = 1 - P(A)$$

一个有用的结果:

$$P(A \cup B) = P(A) + P(B) - P(AB)$$

可以推广为:

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j}\sum P(A_iA_j) + \sum\sum_{i<j<k}\sum P(A_iA_jA_k) + \cdots + (-1)^{n+1}P(A_1\cdots A_n)$$

如果 S 是有限集, 且其中每个结果发生的可能性是一样的, 那么

$$P(A) = \frac{|A|}{|S|}$$

其中 $|E|$ 表示事件 E 所含的结果数.

$P(A)$ 可以理解为频率的趋势或者相信程度的度量.

习　题

1. 一个盒子里有 3 个弹珠, 红、绿、蓝各一个. 先从中取出一个, 再放回, 然后再取出一个, 试描述此样本空间. 如果不放回呢?

2. 连续掷一枚骰子, 直到 6 出现, 试验停止, 试描述此样本空间. 令 E_n 表示“在试验停止时, 一共掷了 n 次”, 那么样本空间里的哪些结果包含在 E_n 里? $\left(\bigcup_{i=1}^{\infty} E_n\right)^c$ 的含义?

3. 掷两枚骰子, 令 E 表示“骰子点数之和为奇数”, 令 F 表示“至少有一个骰子点数为 1”; 令 G 表示“骰子点数之和为 5”. 试描述事件 $EF, E\cup F, FG, EF^c$ 和 EFG.

4. A,B,C 三人轮流掷硬币, 第一次出现正面朝上者为胜. 我们用 0 表示“正面朝下”, 1 表示“正面朝上”, 样本空间可写为:

$$S=\begin{cases}1,01,001,0001,\cdots\\0000\cdots\end{cases}$$

(a) 试解释此样本空间; (b) 表述以下事件:

(i) A 胜了 (记为 A); (ii)B 胜了 (记为 B); (iii) $(A\cup B)^c$.

在该试验中, 假定 A 先掷, B 后掷, 再 C 掷, 接着又是 A 掷……循环往复.

5. 一个系统包含 5 个元件, 每个元件有可能是好的, 也有可能是坏的. 用向量 (x_1,x_2,x_3,x_4,x_5) 表示各个元件的状态, 其中 $x_i=1$ 表示第 i 个元件是好的, $x_i=0$ 表示第 i 个元件是坏的.

(a) 样本空间里一共有多少个结果?

(b) 如果元件 1 和 2 是好的, 或者元件 3 和 4 是好的, 或者元件 1, 3 和 5 都是好的, 那么系统工作正常. 令 W 表示系统工作正常, 写出 W 包含的所有结果.

(c) 令 A 表示元件 4 和 5 都是坏的, 那么 A 里一共有多少结果?

(d) 写出事件 AW 包含的所有结果.

6. 医院管理系统对前来治疗的受枪伤病人进行编号, 其依据为是否买了保险, 如果买了保险, 则记为 1, 否则记为 0; 还根据他们的身体状况, 如果良好, 就记为 g, 如果一般, 就记为 f, 如果严重, 就记为 s.

(a) 给出试验的样本空间;

(b) 令 A 表示“病人病情很严重”, 列出 A 里的所有结果;

(c) 令 B 表示“病人没有买保险”, 列出 B 里的所有结果;

(d) 列出事件 $B^c\cup A$ 里的所有结果.

7. 调查一个业余足球队里 15 名球员的工作, 是蓝领还是白领, 还有政治关系, 是共和党、民主党还是无党派, 问:

(a) 样本空间一共多少结果?

(b) “至少有一个队员是蓝领”有多少结果?

(c) “队员里没有人是无党派人士”有多少结果?

8. 设事件 A 和 B 互不相容, 且 $P(A)=0.3$, $P(B)=0.5$, 求以下事件的概率:

(a) A 或者 B 发生; (b) A 发生但 B 不发生; (c) A 和 B 都发生.

9. 某零售店既接受运通卡也接受维萨卡. 它的顾客中有 24% 的人持有运通卡, 61% 的顾客持有维萨卡, 11% 的顾客持有两种卡, 问至少持有一张卡的人数百分比.

10. 某个学校 60% 的学生既不戴耳环, 也不戴项链, 20% 的学生戴耳环, 30% 的学生戴项链. 如果随机挑一个学生, 求符合以下条件的概率:

(a) 戴耳环或者项链. (b) 既戴耳环也戴项链.

11. 美国男性中有 28% 的人抽烟, 7% 的人抽雪茄, 5% 的人既抽烟也抽雪茄.

(a) 既不抽烟也不抽雪茄的人的百分比是多少?

(b) 只抽雪茄但不抽烟的人的百分比是多少?

12. 某个小学有三个语言班, 一个是西班牙语班, 一个是法语班, 另一个是德语班. 这些语言班对学校里的 100 个学生开放. 其中, 有 28 人参加西班牙语班; 有 26 人参加法语班; 有 16 人参加德语班; 有 12 人既参加西班牙语班也参加法语班; 有 4 人既参加西班牙语班也参加德语班; 有 6 人既参加法语班也参加德语班; 另外, 有 2 人三个班都参加.

(a) 随机选一名学生, 他 (或她) 不参加任何班的概率是多大?

(b) 随机选一名学生, 他 (或她) 恰好参加一个班的概率是多大?

(c) 随机选两名学生, 其中至少有一人参加语言班的概率是多大?

13. 某个人口规模为 100 000 的城镇有三份报纸 I、II 和 III , 以下是对读报人群比例的调查结果:

I: 10% ;　I 和 II: 8%;　I、II 和 III: 1%;

II: 30%;　I 和 III: 2%;　III : 5%;　II 和 III: 4%.

(比如, 有 8 000 人读报纸 I 和 II)

(a) 求出仅仅读一份报纸的人数; (b) 至少读两份报纸的人数;

(c) 如果 I 和 III 是早报, 而 II 是晚报, 那么至少读一份早报和一份晚报的人数为多少?

(d) 有多少人不读报纸? (e) 有多少人仅读一份早报和一份晚报?

14. 对某份杂志的 1 000 名订阅者的调查给出了如下数据: 考虑到他们的工作、婚姻和教育状况, 有 312 名专业人员, 470 名已婚人士, 525 名大学毕业生, 42 名大学毕业的专业人员, 147 名已婚大学毕业生, 86 名已婚专业人员, 25 名已婚且大学毕业的专业人员. 证明这些数据是不正确的.

提示: 令 M, W 和 G 分别表示专业人员、已婚人士及大学毕业生的集合. 假定随机地从这 1 000 人当中选择一人, 利用命题 4.4 来证明: 如果上述数据是正确的, 那么 $P(M \cup W \cup G) > 1$.

15. 从 52 张牌里随机取 5 张, 求以下事件概率:

(a) 同花 (也即 5 张牌同一花色);

(b) 一对 (5 张牌为 a, a, b, c, d 形式 (点数), 其中 a, b, c, d 各不相同);

(c) 两对 (5 张牌为 a, a, b, b, c 形式, 其中 a, b, c 各不相同);

(d) 三张同点数 (5 张牌为 a, a, a, b, c 形式, 其中 a, b, c 各不相同);

(e) 四张同点数 (5 张牌为 a, a, a, a, b 形式, 其中 a, b 不相同).

16. 同时掷 5 枚骰子, 证明:

(a) $P\{\text{每枚都不一样}\} = 0.092\ 6$;　(b) $P\{\text{一对}\} = 0.463\ 0$;

(c) $P\{\text{两对}\} = 0.231\ 5$;　(d) $P\{\text{三面一样}\} = 0.154\ 3$;

(e) $P\{\text{福尔豪斯}\} = 0.038\ 6$;　(f) $P\{4\text{ 枚一样}\} = 0.019\ 3$;

(g) $P\{5\text{ 枚一样}\} = 0.000\ 8$.

17. 如果 8 个车随机地放在国际象棋棋盘上, 求没有一对能互捉的概率, 也即求任何一行或一列至多只有一个车的概率. (国际象棋盘有 $8 \times 8 = 64$ 格, 棋子放在格内)

18. 从一副洗好的扑克牌里随机挑两张, 恰好配成"黑杰克"的概率是多大? (所谓"黑杰克", 也即其中有张"A", 另一张是 10, "J", "Q", "K" 中任一张.)

19. 两个同样的骰子, 各有两面涂成了红色, 两面涂成了蓝色, 一面涂成了黄色, 剩下一面涂成了白色. 同时掷这两枚骰子, 问出现同一种颜色的概率是多大?

20. 假设某赌徒正在和庄家玩二十一点, 对于一副洗好的扑克牌, 赌徒和庄家都分不到"黑杰克"的概率是多大? (每人分配到 2 张牌)

21. 一个小型社区由 20 个家庭组成, 其中只有一个小孩的家庭有 4 个, 有 2 个小孩的家庭有 8 个, 有 3 个小孩的家庭有 5 个, 有 4 个小孩的家庭有 2 个, 有 5 个小孩的家庭有 1 个.

(a) 如果随机选取一个家庭, 它有 i 个孩子的概率是多大? $i = 1, 2, 3, 4, 5$.

(b) 如果随机选取一个孩子, 孩子来自有 i 个孩子的家庭的概率是多大? $i = 1, 2, 3, 4, 5$.

22. 对于 n 张扑克牌, 有一种洗牌技术: 对于任何一副没洗的牌, 考虑第一张, 掷一枚硬币, 如果硬币出现的是正面, 那么这张牌仍留原位, 如果硬币是反面, 那么将这张牌放到最后一张, 接着考虑第二张牌的位置变换, 其规则与第一张牌的相同. 硬币掷了 n 次后, 就完成了一轮洗牌. 比如, 设 $n = 4$, 牌最初的顺序为 1,2,3,4, 如果硬币的顺序为正面、反面、反面、正面, 那么最后牌的顺序为: $1, 4, 2, 3$. 假设所掷硬币是均匀的, 且各次结果相互独立, 问经过一轮洗牌后仍保持原来次序的概率有多大?

23. 同时掷两枚均匀骰子, 问第二枚骰子的点数大于第一枚骰子的点数的概率是多大?

24. 同时掷两枚骰子, 骰子点数之和为 i 的概率是多大? 并求出 $i = 2, 3, \cdots, 11, 12$ 时的值.

25. 同时掷两枚骰子, 直到骰子点数之和为 5 或者 7 出现, 求和为 5 先出现的概率.

提示: 令 E_n 表示第 n 次掷骰子出现和为 5, 但此前 $n-1$ 次当中既不出现和为 5, 也不出现和为 7, 计算 $P(E_n)$ 且证明 $\sum_{n=1}^{\infty} P(E_n)$ 就是题目所求概率.

26. Crap 赌博规则如下: 其中一人先掷两枚骰子, 如果和为 2,3 或者 12, 那么他便输了; 如果和为 7 或者 11, 那么他便赢了. 如果和为其他, 那么由他继续掷骰子, 一直到第一次掷出的和数再次出现, 或者出现和为 7. 若出现的是 7, 则他输了, 若出现的是第一次掷出的和数, 则他赢了. 求他赢的概率.

提示: 令 E_i 表示第一次掷骰子所得点数和为 i, 且取得了胜利. 那么所求概率为 $\sum_{i=2}^{12} P(E_i)$. 为了计算 $P(E_i)$, 定义 $E_{i,n}$ 表示事件"第一次和为 i 且在第 n 次取得了胜利". 证明 $P(E_i) = \sum_{n=1}^{\infty} P(E_{i,n})$.

27. 坛子里有 3 个红球和 7 个黑球, 玩家 A 和 B 从坛子里交替拿球, 直到有人拿到了红球, 求 A 取到了红球的概率. (假设 A 先取球, 然后 B 再取, 接下来又是 A 取, 等等, 并假设球取出来后不放回.)

28. 一个坛子里有 5 个红球、6 个蓝球和 8 个绿球. 如果随机取 3 个球, 问以下事情的概率:

(a) 三个球同一种颜色;　(b) 三个球不同的颜色.

假设取球后, 记下其颜色, 然后再放回坛内 [这就是所谓的*有放回抽样* (Sampling with relacement)], 重新计算以上事件的概率.

29. 坛子里有 n 个白球和 m 个黑球, 其中 n 和 m 都是正数.

(a) 随机取两个球, 问它们是同一种颜色的概率?

(b) 从坛子里随机取一个球, 然后放回再第二次取球, 求取出的两个球颜色相同的概率.

(c) 证明 (b) 的概率始终大于 (a) 的概率.

30. 两所学校的棋类俱乐部分别有 8 和 9 名棋手, 每个俱乐部各随机选 4 名参加两校间的对抗赛. 选出来的棋手随机地和另一俱乐部选出来的棋手进行两两配对下棋, 假设丽贝卡和妹妹埃莉斯分别在这两校的棋类俱乐部, 求以下事件的概率:

(a) 丽贝卡和埃莉斯正好配成一对. (b) 丽贝卡和埃莉斯都被选出, 但是她们不下棋;

(c) 她俩人只有一个被选出代表学校参赛.

31. 一个 3 人篮球队包括 1 个后卫, 1 个前锋, 1 个中锋.

(a) 如果从 3 个这样的篮球队里分别选一人, 正好可以组成一个新篮球队的概率是多大?

(b) 选出来的 3 人是打同一位置的概率是多大?

32. 一个小组有 b 个男孩, g 个女孩, 按随意顺序站成一排, 即 $(b+g)!$ 种排列中任一种都是等可能的, 试问, 第 i 个位置 $(1 \leqslant i \leqslant b+g)$ 站的正好是女孩的概率.

33. 树林里有 20 只麋鹿, 捉住其中 5 只, 贴上标签, 然后再放回. 一段时间后, 再捉住 4 只麋鹿. 其中有两只贴了标签的概率是多大? 此处做了什么样的假设?

34. 据报道, Yarborough 二世曾经用 1000 比 1 的赌注打赌一个桥牌手里的 13 张牌里至少有一张牌是 10 或者更大 (所谓更大意味着是 10, 或者“J”,“Q”,“K”, 和“A”). 如今, 一手牌里如果没有一张 10 以上的牌, 就称为 Yarborough. 问随机发的一手牌是 Yarborough 的概率是多大?

35. 一个坛子装有 12 个红球, 16 个蓝球, 18 个绿球. 从中随机地取出 7 个球. 求出下列事件的概率.

(a) 抽出 3 个红球, 2 个蓝球, 2 个绿球;

(b) 其中至少有 2 个红球;

(c) 抽出的球的颜色全相同;

(d) 抽出的球中恰有 3 个红球或者恰有 3 个蓝球.

36. 从 52 张牌里随机取 2 张, 求以下事件概率:

(a) 都是“A”; (b) 点数相同.

37. 老师给班里学生留了 10 道习题, 并且告诉他们期末考试就是从中随机选择 5 道, 如果有个学生解出了其中 7 道, 求以下事件概率:

(a) 他 (她) 能解出所有的 5 道考试题; (b)至少能解出其中 4 道.

38. 抽屉里有 n 只袜子, 其中 3 只红的. 如果随机取两只袜子, 同为红色的概率为 1/2, 试问 n 为多大?

39. 城镇里有 5 个旅馆, 某天有 3 人入住旅馆, 问正好住进不同旅馆的概率是多大? 其中做了什么样的假设?

40. 城镇里有 4 人修理电视机, 现在有 4 台坏电视机, 问正好有 $i(i=1,2,3,4)$ 人被要求参与修理的概率? 其中做了什么样的假设?

41. 掷一枚骰子 4 次, 至少出现一次 6 的概率是多大?

42. 连续掷两枚骰子 n 次, 计算双 6 至少出现一次的概率. 要想此概率大于或等于 1/2, n 至少要多大?

43. (a) 包含 A 和 B 在内的 N 个人随机的排成一排, 问 A 和 B 紧挨着的概率是多大;

(b) 如果是随机的排成一圈, 那么这个概率是多大?

44. 有 A,B,C,D,E 五人, 站成一排, 假设每种顺序都是等可能的, 求以下事件概率:

(a) A 和 B 之间恰好有一个人; (b) A 和 B 之间恰好有两个人;

(c) A 和 B 之间恰好有三个人.

45. 有位妇女有 n 把钥匙, 其中有一把能打开房门.

(a) 她随机地用钥匙开房门, 如果打不开, 就换一把钥匙 (钥匙不重复试用), 请问正好在第 k 次成功打开房门的概率是多大?

(b) 如果钥匙可重复试用, 这时 (a) 中问题的概率是多大?

46. 需要多少人, 才能保证其中至少有两人同一月份过生日的概率大于 1/2? 假设每个月的可能性是一样的.

47. 房间里有 12 个人, 求没有两人在同一月份出生的概率.

48. 有 20 人, 求如下事件的概率: 12 个月当中, 其中 4 个月每月均有 2 人过生日, 而另有 4 个月每月均有 3 人过生日.

49. 6 个男人、6 个女人随机地分成 2 组, 每组 6 人, 求两组的男人数正好一样的概率.

50. 在桥牌比赛中, 计算你有 5 张黑桃, 而搭档正好有 8 张黑桃的概率.

51. n 个球随机地放到 N 个房间, 设 N^n 种结果每种都是等可能的, 试求第一个房间恰有 m 个球的概率.

52. 衣柜里有 10 双鞋, 随机拿 8 只, 问如下事件的概率:

(a) 一双鞋都没有; (b)正好有一双鞋.

53. 如果 4 对夫妇坐成一排, 试求没有夫妇坐在一起的概率.

54. 计算桥牌比赛中有一家至少缺一套花色的概率. 此答案并不是 $\binom{4}{1}\binom{39}{13}\Big/\binom{52}{13}$, 为什么?

提示: 利用命题 4.4.

55. 一手牌 13 张, 求以下事件概率: (a) 有同花色的“A”和“K”; (b) 有同点数的四张.

56. 有两人玩以下游戏: A 从图 2.6 中所示的三个轮盘中选择一个, 然后 B 在剩下的两个中选择一个. 接着两人分别转动各自选中的轮盘, 轮盘最后所停的位置下方区域里的数字大者获胜. 假定每个轮盘停在三个区域是等可能的. 如果是你, 你会选择是 A 还是 B? 并解释原因.

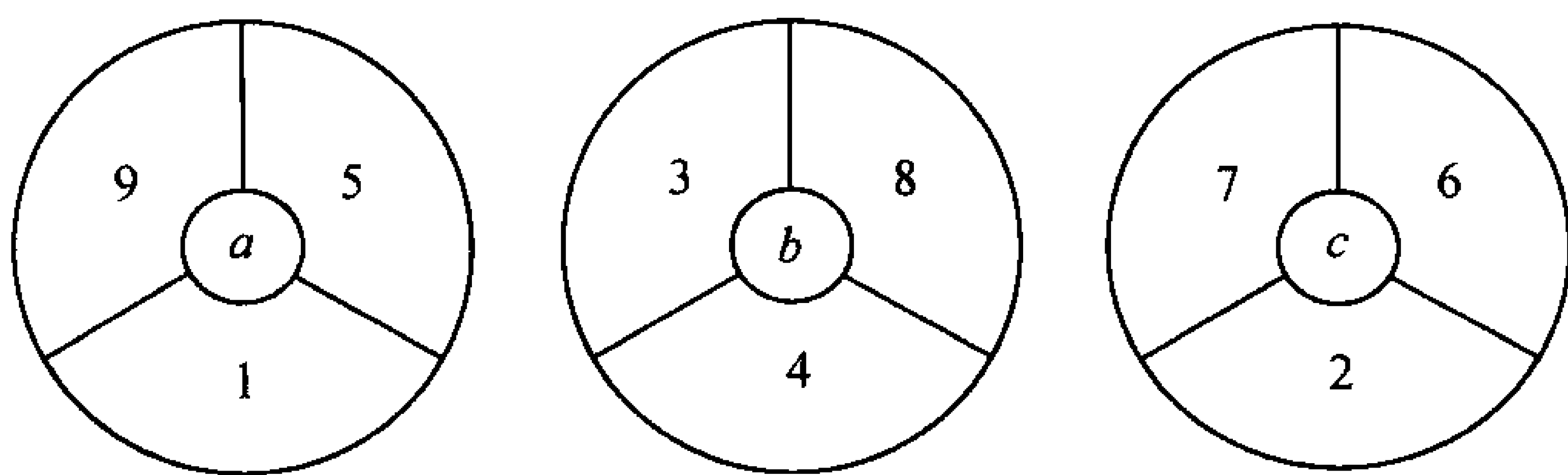


图 2.6 习题 56 图

理论习题

1. 证明: $EF \subset E \subset E \cup F$.

2. 证明: 如果 $E \subset F$, 那么 $F^c \subset E^c$.

3. 证明: $F = FE \cup FE^c$ 和 $E \cup F = E \cup E^cF$.

4. 证明: $\left(\bigcup_{i=1}^{\infty} E_i\right)F = \bigcup_{i=1}^{\infty} E_iF$ 以及 $\left(\bigcap_{i=1}^{\infty} E_i\right) \cup F = \bigcap_{i=1}^{\infty}(E_i \cup F)$.

5. 对任一系列事件 $E_1, E_2, \cdots$, 定义一列两两不相交事件 $F_1, F_2, \cdots$(如 $i \neq j$ 则 $F_iF_j = \varnothing$) 使得对任何 $n \geqslant 1$ 有

$$\bigcup_{i=1}^{n} F_i = \bigcup_{i=1}^{n} E_i$$

6. 设 E, F, G 为三个事件, 写出如下事件的表达式:

(a) 仅仅 E 发生; (b) E 和 G 都发生, 但 F 不发生; (c) 至少有一个事件发生;
(d) 至少有两个事件发生; (e) 三个事件都发生; (f) 一个事件都不发生;
(g) 最多一个事件发生; (h) 最多两个事件发生; (i) 正好两个事件发生;
(j) 最多三个事件发生.

7. 化简下列表达式:

(a) $(E \cup F)(E \cup F^c)$; (b) $(E \cup F)(E^c \cup F)(E \cup F^c)$; (c) $(E \cup F)(F \cup G)$.

8. 给定集合 S, 如果存在一列互斥非空子集 $S_1, S_2, \cdots, S_k (k > 0)$, 满足 $\bigcup_{i=1}^{k} S_i = S$, 那么称 $\{S_1, S_2, \cdots, S_k\}$ 为 S 的一个分割(partition). 令 T_n 表示集合 $\{1, 2, \cdots, n\}$ 的不同分割的总数, 因此, $T_1 = 1$(表示仅仅一个分割 $S_1 = \{1\}$), $T_2 = 2$(表示两个分割: $\{\{1,2\}\}$ 和 $\{\{1\},\{2\}\}$).

(a) 证明: $T_3 = 5, T_4 = 15$.

(b) 证明:

$$T_{n+1} = 1 + \sum_{k=1}^{n} \binom{n}{k} T_k$$

并利用此公式计算 T_{10}.

提示: 在 $(n+1)$ 个元素中选定一个元素, 称为特殊元素. 并将分割分类: 分割中含有这个特殊元素的子集称为这个分割的特殊子集. 具有相同特殊子集的分割形成一子类, 再将每一个子类所含分割数加起来, 就得到 T_{n+1}.

9. 设一个试验的样本空间为 S, 将试验重复 n 次, 对样本空间 S 里的任一事件 E, 令 $n(E)$ 表示 n 次事件中 E 发生的次数, 定义 $f(E) = n(E)/n$. 证明 $f(\cdot)$ 满足公理 1, 2, 3.

10. 证明: $P(E \cup F \cup G) = P(E) + P(F) + P(G) - P(E^cFG) - P(EF^cG) - P(EFG^c) - 2P(EFG)$.

11. 如果 $P(E) = 0.9, P(F) = 0.8$, 证明 $P(EF) \geqslant 0.7$. 更一般地, 证明 Bonferroni 不等式:

$$P(EF) \geqslant P(E) + P(F) - 1$$

12. 证明: E 和 F 恰好只有一个发生的概率为 $P(E) + P(F) - 2P(EF)$.

13. 证明: $P(EF^c) = P(E) - P(EF)$.

14. 用数学归纳法证明命题 4.4.

15. 一个坛子里有 M 个白球和 N 个黑球, 随机取 r 个, 问恰好取到 k 个白球的概率是多少?

16. 将 Bonferroni 不等式推广到 n 个事件的情形, 也即证明:

$$P(E_1E_2\cdots E_n) \geqslant P(E_1) + \cdots + P(E_n) - (n-1)$$

17. 考虑例 5m 中的匹配问题. 令 A_N 表示 N 个人都不选自己帽子的方法数, 指出

$$A_N = (N-1)(A_{N-1} + A_{N-2})$$

结合边界条件 $A_1 = 0, A_2 = 1$ 可以递推地求出 A_N. 最后, 没有人拿到自己帽子的概率为 $A_N/N!$.

提示: 可将 N 个人都不选自己帽子的方法分类, 指定其中一个人, 称为“老王”, 老王选定了一顶别人的帽子, 这样就把选帽子的方法按老王选定的帽子的不同分成 $(N-1)$ 类, 设老王选定了另一个人 (老张) 的帽子, 现在有一个特殊的人 (老张, 他已经没有自己的帽子可选), 以及一个特殊的帽子 (老王的帽子), 再将剩下的 $N-1$ 人都不选自己帽子的方法分成两类, 一类是特殊人选了特殊的帽子, 一类是特殊人不选特殊的帽子, 分别计算两类中的方法数, 就可以导出所得的公式.

18. 令 f_n 表示掷一枚硬币 n 次且从不出现连续正面的可能结果数, 证明:

$$f_n = f_{n-1} + f_{n-2} \quad n \geqslant 2, \text{其中} f_0 \equiv 1, f_1 \equiv 2$$

提示: 按第一次掷硬币的结果将可能的试验结果分成两类, 一类是反面朝上, 另一类是正面朝上, 分别对两类结果计数. 若 P_n 表示 n 次掷硬币的结果中不会连续出现正面的概率, 在掷 n 枚硬币的各种结果为等可能情况下, 找出 P_n 与 f_n 之关系, 并计算 P_{10}.

19. 一个坛子里有 n 个红球, m 个蓝球, 从中一个一个取球, 一直到取了 $r(r \leqslant n)$ 个红球为止. 求此时正好取出 k 个球的概率是多少?

提示: 正好取出 k 个球等价于第 k 次取出红球且前 $k-1$ 次取出 $r-1$ 个红球.

20. 设有一个试验, 其样本空间包含可数个结果, 试证明不可能所有可能结果发生的概率都一样. 有没有这样的可能性: 所有的可能结果发生的概率均为正数?

***21.** 在例 5o 中, 讨论了在 n 次赢和 m 次输的次序是随机的情况下, 关于赢的游程的数目的概率计算问题. 现在考虑全部游程的个数 (赢的游程的数目加上输的游程的数目), 证明:

$$P\{2k\text{个游程}\} = 2\frac{\binom{m-1}{k-1}\binom{n-1}{k-1}}{\binom{m+n}{n}}$$

$$P\{2k+1\text{个游程}\} = \frac{\binom{m-1}{k-1}\binom{n-1}{k} + \binom{m-1}{k}\binom{n-1}{k-1}}{\binom{m+n}{n}}$$

自检习题

1. 咖啡馆有主菜、主食和甜点三道供餐, 可能的选择如下:

种 类	选 择
主菜	鸡肉或烤牛肉
主食	面、米饭或土豆
甜点	冰淇淋、果冻、苹果酱或桃子

客人在每个种类中选择一种.

(a) 样本空间里一共有多少个结果?

(b) 令 A 表示“选择冰淇淋”, A 里一共有多少结果?

(c) 令 B 表示“选了鸡肉”, B 里一共多少结果?

(d) 列举事件 AB 里所有结果;

(e) 令 C 表示“选了米饭”, C 里一共有多少结果?

(f) 列举事件 ABC 所含的所有结果.

2. 时装店里来了一位顾客, 已知他买西装的概率为 0.22, 买衬衫的概率为 0.30, 买领带的概率为 0.28, 既买西装又买衬衫的概率为 0.11, 既买西装又买领带的概率为 0.14, 既买衬衫又买领带的概率为 0.10, 三者都买的概率为 0.06. 求以下事件的概率:

(a) 一样都不买; (b)正好买一样.

3. 随机发一副牌, 第 14 张是“A”的概率是多少? 第 14 张才出现“A”的概率又是多少?

4. 令 A 表示“洛杉矶的城中气温为 70°F”, 令 B 表示“纽约的城中气温为 70°F”, 再令 C 表示“洛杉矶和纽约的城中气温较高者为 70°F”. 如果 $P(A) = 0.3, P(B) = 0.4$, 且 $P(C) = 0.2$, 求“洛杉矶和纽约的城中气温较低者为 70°F”发生的概率.

5. 一副洗好的牌共 52 张, 求最上面 4 张是以下情况的概率: (a) 不同点数; (b) 不同花色.

6. 坛子 A 里有 3 个红球和 3 个黑球, 而坛子 B 里有 4 个红球和 6 个黑球, 从两个坛子里各取一球, 正好是同一种颜色的概率是多少?

7. 某个州发行一种彩票, 彩民要从 1 到 40 个数里选 8 个. 最后, 组委会也从这 40 个数里选 8 个作为中奖数字, 假定 $\binom{40}{8}$ 种结果都是等可能的, 求以下事件的概率:

(a) 彩民猜中了 8 个数字; (b) 彩民猜中了 7 个数字; (c)彩民至少猜中了 6 个数字.

8. 从 3 名一年级新生、4 名二年级学生、5 名三年级学生、3 名毕业班学生里随机选择 4 人组成委员会, 求以下事件概率:

(a) 委员会中每个年级恰好一个人;

(b) 委员会由两个二年级学生和两个三年级学生组成;

(c) 委员会仅由二年级或三年级学生组成.

9. 对于有限集 A, 令 $N(A)$ 表示集合 A 里元素的个数.

(a) 证明: $N(A \cup B) = N(A) + N(B) - N(AB)$

(b) 更一般地, 证明:

$$N\left(\bigcup_{i=1}^{n} A_i\right) = \sum_i N(A_i) - \sum_{i<j}\sum N(A_iA_j) + \cdots + (-1)^{n+1}N(A_1\cdots A_n)$$

10. 赛马比赛中有 6 匹马, 编为 1, 2, 3, 4, 5, 6 号. 样本空间由 6! 种可能的比赛结果组成. 令 A 表示“1 号马跑在前 3 名”, 令 B 表示“2 号马跑第二名”, 那么 $A \cup B$ 一共包含多少个结果?

11. 从一副洗好的 52 张牌里取 5 张, 每个花色至少有一张的概率是多大?

12. 篮球队有 6 名前场队员和 4 名后场队员, 现在要随机两两配对分宿舍. 问正好有两间宿舍由前场队员和后场队员合住的概率是多大?

13. 某人从“R E S E R V E”中随机挑选一个字母, 再从“V E R T I C A L”中选一个, 求两个字母恰好相同的概率.

14. 证明布尔不等式：

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) \leqslant \sum_{i=1}^{\infty} P(A_i).$$

15. 证明：如果 $P(A_i)=1$ 对任一 $i \geqslant 1$ 成立，那么 $P\left(\bigcap_{i=1}^{\infty} A_i\right)=1$.

16. 令 $T_k(n)$ 表示"集合 $\{1,\cdots,n\}$ 分成 k 个非空子集的不同分割数"，此处 $1 \leqslant k \leqslant n$(分割的定义参见理论习题 8)，证明：$T_k(n)=kT_k(n-1)+T_{k-1}(n-1)$

提示：$\{1\}$ 是子集的分割数是多少？$\{1\}$ 不是子集的分割数是多少？

17. 一个坛子里装有 5 个红球，6 个白球和 7 个蓝球，无放回地随机抽取 5 个，三种颜色的球都取到的概率是多大？

18. 将 4 个红球，8 个蓝球和 5 个绿球随机地排成一行.

(a) 前 5 个球为蓝色的概率有多大？

(b) 前 5 个球中没有蓝色球的概率有多大？

(c) 最后 3 个球为三种不同颜色的概率有多大？

(d) 所有红球连在一起的概率有多大？

19. 在一副 52 张扑克牌中随机地抽出 10 张牌. 将抽出的牌按其花色分成四堆.

(a) 各堆中张数分别为 4, 3, 2, 1 的概率有多大？

(b) 其中两堆各有 3 张牌，还有一堆有 4 张牌的概率有多大？

20. 设一坛子中有 20 个红球，10 个蓝球. 现在随机地一个一个地从坛子中取出. 求红球比蓝球先取完的概率.

第 3 章　条件概率和独立性

3.1　简　　介

本章, 我们将介绍概率论中最重要的概念之一: 条件概率. 此概念的重要性在于两方面. 一方面, 我们在计算某些事件的概率时, 同时具有某些关于该试验的附加信息, 此时概率应该是条件概率. 另一方面, 即使手头没有附加信息, 也可以利用条件概率的方法计算某些事件的概率, 而这种方法常使计算变得十分简单.

3.2　条件概率

同时掷两枚骰子, 并假设 36 种结果都是等可能发生的, 因此每种结果发生的概率为 1/36. 进一步假设已知第一枚骰子点数为 3, 在这些条件下两枚骰子点数之和为 8 的概率是多大? 为了计算这个概率, 解释如下: 既然第一枚骰子点数为 3, 那么掷两枚骰子一共有 6 种可能结果: $(3,1),(3,2),(3,3),(3,4),(3,5),(3,6)$. 因为每个结果发生的概率都一样, 那么, 这 6 种结果也应该是等可能的, 即在给定第一个骰子为 3 的情况下, 下面 6 种结果: $(3,1),\cdots,(3,6)$, 每一个结果发生的 (条件) 概率应该是 1/6, 而样本空间中其他 30 个点的 (条件) 概率应该是 0. 这样, 在第一枚骰子点数为 3 的条件下, 两枚骰子点数之和为 8 的概率应该是 1/6.

如果令 E 和 F 分别表示"两枚骰子点数之和为 8"和"第一个骰子点数为 3", 利用上述方法, 计算得到的概率称为假定 F 发生的情况下 E 发生的条件概率, 记为

$$P(E|F)$$

用如下方式可以推导出一个对于所有 E 和 F 都适用的计算 $P(E|F)$ 的常用公式: 如果 F 发生了, 那么为了 E 发生, 其结果必然是既属于 E 也属于 F, 即这个结果必然属于 EF. 既然已知 F 已经发生, F 成了新的样本空间, 因此 E 发生的 (条件) 概率必然等于 EF 发生的概率与 F 发生的概率之比值. 因此, 有如下定义.

定义　如果 $P(F)>0$, 那么

$$P(E|F)=\frac{P(EF)}{P(F)} \tag{2.1}$$

例 2a 某个学生参加一个时限为 1 小时的测验. 假定对任意 $0 \leqslant x \leqslant 1$ 来说, 他在 x 小时内完成测验的概率为 $x/2$, 已知 0.75 小时后他仍在答题, 问他最后要用光一小时的条件概率是多少?

解: 令 L_x 表示"学生完成测验的时间不超过 x 小时 $(0 \leqslant x \leqslant 1)$", 令 F 表示"学生用光了 1 小时". 因为事件 F 就是"学生完成测验时间超过 1 小时", 因此

$$P(F) = P(L_1^c) = 1 - P(L_1) = 0.5$$

因为事件"0.75 小时后学生仍在答题"就是事件"学生完成测验时间超过 0.75 小时", 可记为 $L_{0.75}^c$, 因此所求概率为

$$P(F|L_{0.75}^c) = \frac{P(FL_{0.75}^c)}{P(L_{0.75}^c)} = \frac{P(F)}{1 - P(L_{0.75})} = \frac{0.5}{0.625} = 0.8$$ ■

若有限样本空间 S 中, 每一个试验结果是等可能的, 那么当以 S 的子集 F 作为条件时, 所有 F 中的结果也都是等可能的. 因此在计算条件概率时, 可将 F 作为样本空间, 用压缩了的样本空间通常使计算变得简单, 概念也容易理解. 下面几个例子就说明了这点.

例 2b 抛掷一枚硬币两次, 假定样本空间 S={(H,H),(H,T),(T,H),(T,T)} 中的四个样本点发生的可能性是一样的, 求给定以下事件后两枚硬币都是正面朝上的条件概率: (a) 第一枚正面朝上. (b) 至少有一枚正面朝上.

解: 令 B={(H,H)} 表示事件"两枚硬币都是正面朝上"; 令 F={(H,H),(H,T)} 表示事件"第一枚硬币正面朝上"; 令 A={(H,H),(H,T),(T,H)} 表示事件"至少有一枚硬币正面朝上". 那么 (a) 的所求概率为

$$P(B|F) = \frac{P(BF)}{P(F)} = \frac{P(\{(\mathrm{H},\mathrm{H})\})}{P(\{(\mathrm{H},\mathrm{H}),(\mathrm{H},\mathrm{T})\})} = \frac{1/4}{2/4} = 1/2$$

对于 (b), 有

$$P(B|A) = \frac{P(BA)}{P(A)} = \frac{P(\{(\mathrm{H},\mathrm{H})\})}{P(\{(\mathrm{H},\mathrm{H}),(\mathrm{H},\mathrm{T}),(\mathrm{T},\mathrm{H})\})} = \frac{1/4}{3/4} = 1/3$$

因此, 已知第一枚硬币正面朝上的条件下, 两枚硬币都是正面朝上的条件概率为 1/2, 而已知至少一枚硬币正面朝上的条件下, 两枚硬币都是正面朝上的概率为 1/3. 很多学生对后者感到吃惊, 他们认为至少有一枚正面朝上这个事件有两种可能性, 两枚都正面朝上和只有一枚正面朝上, 他们的错误是把这两种情况看成等可能的了. 事实上, "至少一面朝上"这个事件包含了 3 个结果: (H,H),(H,T),(T,H). 而这三种结果都是等可能的, 而 (H,H) 只是其中一个结果, 因此其条件概率为 1/3 是很自然的了. ■

例 2c 桥牌游戏里, 52 张牌平均发给东、西、南、北四家. 如果南和北一共有 8 张黑桃, 问: 东有剩下 5 张黑桃里的 3 张的概率是多大?

解: 最简单的方法是缩减样本空间. 也即, 已知南北 26 张牌中共有 8 张黑桃,

那么还剩下 26 张牌, 其中正好 5 张是黑桃, 将要分给东西两家. 由于每种分法都是等可能的, 因此, 东家 13 张牌中正好有 3 张黑桃的条件概率是

$$\binom{5}{3}\binom{21}{10}\Big/\binom{26}{13} \approx 0.339 \qquad ■$$

例 2d　一个坛子里有 r 个红球和 b 个蓝球, 随机地从中无放回地依次取出 n 个球 $(n \leqslant r+b)$. 已知其中 k 个是蓝球, 问第一个球是蓝球的条件概率是多大?

解: 如果我们设想所有的球都标了号, 其中蓝球标有 1 到 b 号, 红球标有 $b+1$ 到 $b+r$ 号. 那么从 n 个球中无放回地取 n 个球的结果就对应一个分量各不相同的向量 $x_1, \cdots, x_n$, 其中任一 x_i 的取值介于 1 到 $r+b$ 之间. 假设结果为任何一个向量的可能性是一样的. 因此, 已知向量包含 k 个蓝球 (也即, 包含了 1 到 b 之间的 k 个值) 条件下, 各个结果都是等可能的. 在选中的 n 个球中 (其中 k 个球为蓝球), 由于任何一个球被第一次选中的可能性都是一样的. 因此, 所求概率为 k/n.

如果我们不选择缩减样本空间, 可以这样解决: 令 B 表示"第一次取球为蓝球"这一事件, 而 B_k 为"n 个球中含 k 个蓝球"这一事件, 那么

$$P(B|B_k) = \frac{P(BB_k)}{P(B_k)} = \frac{P(B_k|B)P(B)}{P(B_k)}$$

现在, $P(B_k|B)$ 为如下事件概率: 随机从装有 r 个红球, $b-1$ 个蓝球的坛子里随机抽取 $n-1$ 个球, 正好有 $k-1$ 个蓝球. 这样,

$$P(B_k|B) = \binom{b-1}{k-1}\binom{r}{n-k}\Big/\binom{r+b-1}{n-1}$$

利用上述结论, 以及 $P(B) = \dfrac{b}{r+b}$ 和超几何概率 $P(B_k) = \binom{b}{k}\binom{r}{n-k}\Big/\binom{r+b}{n}$, 可得结果 $P(B|B_k) = k/n$. ■

在 (2.1) 式两边同时乘以 $P(F)$, 可以得到

$$P(EF) = P(F)P(E|F) \tag{2.2}$$

也就是说, 公式 (2.2) 说明了 E 和 F 同时发生的概率, 等于 F 发生的概率乘以 F 发生的条件下 E 发生的条件概率. 公式 (2.2) 在计算事件的交的概率时非常有用.

例 2e　为选修法语课还是选修化学课这件事, 茜琳犹豫不决. 她估计如果选修法语课, 则获得"A"等成绩的概率为 1/2, 而如果选修化学课, 则获得"A"等成绩的概率为 2/3. 如果茜琳通过掷硬币来作决定, 问她将选修化学课, 并获得"A"等成绩的概率是多大?

解: 如果令 C 表示"茜琳选修化学课", 而 A 表示"她获得了'A'等成绩", 那么所求概率为 $P(CA)$, 利用公式 (2.2) 计算如下:

$$P(CA) = P(C)P(A|C) = \frac{1}{2} \times \frac{2}{3} = \frac{1}{3} \qquad ■$$

例 2f　坛子中有 8 个红球与 4 个白球, 现在顺序地无放回地从坛子中取出两

个球. (a) 假定每次抽取时, 坛子中各球被取中的可能性是一样的, 问取出的两个球都是红球的概率是多大? (b) 假定坛子中各个球有不同的权数, 每个红球的权数为 r, 每个白球的权数为 w. 坛子中一个球被抽出的概率等于这个球的权数与当时坛子中全体球的权数之和的比值. 现在问顺序抽出的两个球都是红色的概率有多大?

解: 令 R_1 和 R_2 分别表示第一次与第二次取出红球的事件. 若取出的第一个球是红球, 那么坛子中剩下 7 个红球 4 个白球, 因此, $P(R_2|R_1)=7/11$. 由于 $P(R_1)$ 显然等于 8/12, 所求概率为 $P(R_1R_2)=P(R_1)P(R_2|R_1)=(2/3)(7/11)=14/33$. 当然, 这个概率也可由公式 $P(R_1R_2)=\binom{8}{2}\Big/\binom{12}{2}$ 算出.

至于 (b), 我们还是令 R_i 为第 i 次抽得红球这一事件, 并利用公式

$$P(R_1R_2)=P(R_1)P(R_2|R_1)$$

进行计算. 现在将红球进行标号, 并且记 B_i 为"第一次抽得红球 i"这一事件, 这样

$$P(R_1)=P\left(\bigcup_{i=1}^{8}B_i\right)=\sum_{i=1}^{8}P(B_i)=\frac{8r}{8r+4w}.$$

再者, 当第一次抽得红球以后, 坛子里只剩下 7 个红球和 4 个白球. 与前面的论证相似, 我们得到

$$P(R_2/R_1)=\frac{7r}{7r+4w}$$

这样, 两次抽得红球的概率是

$$P(R_1R_2)=\frac{8r}{8r+4w}\cdot\frac{7r}{7r+4w}.$$

■

公式 (2.2) 的推广有时也称为*乘法规则*, 它提供了任意个事件交的概率的计算方法.

乘法规则

$$P(E_1E_2E_3\cdots E_n)=P(E_1)P(E_2|E_1)P(E_3|E_1E_2)\cdots P(E_n|E_1\cdots E_{n-1})$$

为了证明乘法规则, 对等式右边利用条件概率的定义, 可得

$$P(E_1)\frac{P(E_1E_2)}{P(E_1)}\frac{P(E_1E_2E_3)}{P(E_1E_2)}\cdots\frac{P(E_1E_2\cdots E_n)}{P(E_1E_2\cdots E_{n-1})}=P(E_1E_2\cdots E_n)$$

例 2g 在第 2 章例 5m 的"配对问题"中已经指出, N 个人都没有拿到自己的帽子的概率 P_N 由下式给出

$$P_N=\sum_{i=0}^{N}(-1)^i/i!$$

现在的问题是: N 个人中恰有 k 个人拿到自己的帽子的概率是多少?

解: 现在在 N 个人中选定 k 个人, 求出这 k 个人拿到自己的帽子但其他人都

没有拿到自己的帽子的概率. 记 E 为事件"这 k 个人拿到自己的帽子", G 为事件"其余 $N-k$ 个人都没有拿到自己的帽子". 利用条件概率的定义,

$$P(EG)=P(E)P(G|E)$$

现在记 F_i 为事件"这个子集合中第 i 个人拿到自己的帽子", $i=1,\cdots,k$. 利用乘法规则得到

$$\begin{aligned}P(E)&=P(F_1F_2\cdots F_k)\\&=P(F_1)P(F_2|F_1)P(F_3|F_1F_2)\cdots P(F_k|F_1\cdots F_{k-1})\\&=\frac{1}{N}\cdot\frac{1}{N-1}\cdot\frac{1}{N-2}\cdots\frac{1}{N-k+1}=\frac{(N-k)!}{N!}\end{aligned}$$

已知这 k 个人拿到自己的帽子的条件下, 其余 $N-k$ 个人就随机地从这 $N-k$ 个帽子中选择帽子. 这样, 问题就转化为 $N-k$ 个人都没有拿到自己的帽子的计算问题. 因此

$$P(G|E)=P_{N-k}=\sum_{i=0}^{N-k}(-1)^i/i!$$

这样, k 个特定的人拿到自己的帽子, 而其余的人没有拿到自己的帽子的概率为

$$P(EG)=\frac{(N-k)!}{N!}P_{N-k}$$

而"恰有 k 个人拿到自己的帽子"的事件发生的充要条件是"其中某 k 个人拿到自己的帽子, 但其余的人都没有拿到自己的帽子"的发生. 这样

$$\begin{aligned}P(\text{恰有 } k \text{ 个人拿到自己的帽子})&=\binom{N}{k}P(G|E)=P_{N-k}/k!\\&\approx \mathrm{e}^{-1}/k! \quad (N\text{ 很大时})\end{aligned}$$

■

现在我们利用乘法规则得到第 2 章的例 5h(b) 的第二种解法.

例 2h 一副 52 张牌随机地分成 4 堆, 每堆 13 张. 计算每一堆正好有一张"A"的概率.

解: 定义事件 $E_i, i=1,2,3,4$ 如下:

$E_1=\{$黑桃"A"在任何一堆里$\}$

$E_2=\{$黑桃"A"和红桃"A"在不同的堆里$\}$

$E_3=\{$黑桃"A", 红桃"A"和方块"A"在不同的堆里$\}$

$E_4=\{$4 张"A"在不同的堆里$\}$

所求概率为 $P(E_1E_2E_3E_4)$, 利用乘法规则,

$$P(E_1E_2E_3E_4)=P(E_1)P(E_2|E_1)P(E_3|E_1E_2)P(E_4|E_1E_2E_3)$$

由于 E_1 为样本空间 S, $P(E_1)=1$, 红桃"A"可以在黑桃"A"这一堆, 也可分在其

余 3 堆中. 红桃"A"分在其余 3 堆的可能性为 $P(E_2|E_1) = 39/51$. 方块"A"可以分在黑桃"A"或红桃"A"这些堆里, 也可分在其余的两堆里. 分到其余两堆的可能性为 $P(E_3|E_1E_2) = 26/50$. 梅花"A"可以分在黑桃"A", 红桃"A"或方块"A"这些堆里, 或者分在另外一堆里. 方块"A"在另外一堆的概率为 $P(E_4|E_1E_2E_3) = 13/49$. 因此, 我们可以得到所求的每堆恰好有 1 个"A"的概率为

$$P(E_1E_2E_3E_4) = \frac{39 \times 26 \times 13}{51 \times 50 \times 49} \approx 0.105$$

也即, 大概有 10.5% 的机会每堆牌中恰好有一个"A"(习题 13 将利用乘法规则给出另一种解法). ■

注释 $P(E|F)$ 的定义与概率的频率解释是一致的. 假设我们进行了 n 次独立重复试验 (n 相当大). 若我们只考虑事件 F 发生的那些试验, 此时 $P(E|F)$ 近似地等于事件 E 发生的相对频率. 由于概率 $P(F)$ 是事件 F 发生的频率的极限, 在 n 次独立重复试验中, 事件 F 会近似地发生 $nP(F)$ 次. 类似地, 事件 EF 会近似地发生 $nP(EF)$ 次. 这样, 在 F 发生的近 $nP(F)$ 次试验中, 事件 E 也发生的相对频率近似地等于 $nP(EF)/nP(F) = P(EF)/P(F)$. 当 n 越来越大时, 其相对频率趋于 $P(EF)/P(F)$. 这个值也就是 $P(E|F)$ 的频率定义. 因此, 从概率的频率解释来看 (2.1) 式关于条件概率的定义也是合理的.

3.3 贝叶斯公式

令 E 和 F 都表示事件. 可以将 E 表示为 $E = EF \cup EF^c$ 这样, E 里的结果, 要么同时属于 E 和 F, 要么只属于 E 但不属于 F(见图 3.1). 显然, EF 和 EF^c 是互不相容的, 因此, 根据公理 3 可得

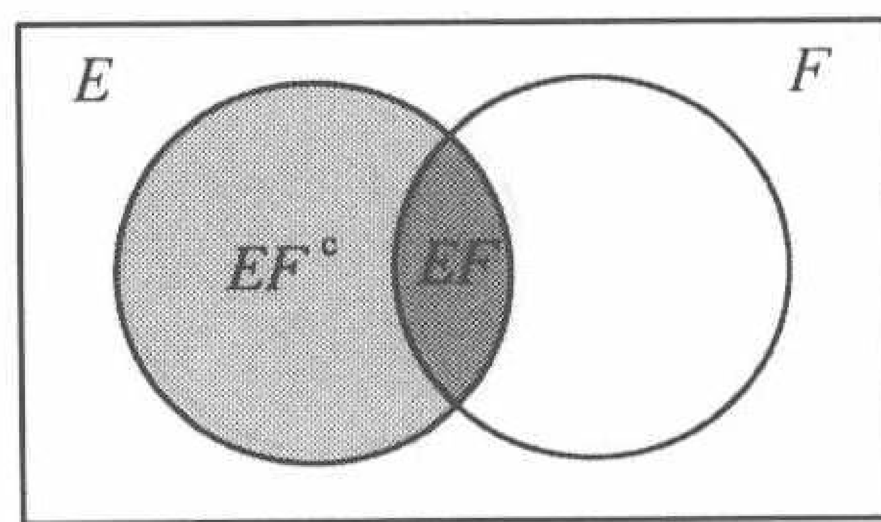


图 3.1 $E = EF \cup EF^c$. EF = 深灰色区域; EF^c = 浅灰色区域

$$\begin{aligned} P(E) &= P(EF) + P(EF^c) \\ &= P(E|F)P(F) + P(E|F^c)P(F^c) \\ &= P(E|F)P(F) + P(E|F^c)[1 - P(F)] \end{aligned} \tag{3.1}$$

公式 (3.1) 说明了事件 E 发生的概率, 等于在 F 发生的条件下 E 的条件概率与在 F 不发生的条件下 E 发生的条件概率的加权平均, 其中加在每个条件概率上的权重就是作为条件的事件发生的概率. 这是一个非常有用的公式, 它使得我们能够通过以第二个事件发生与否作为条件来计算第一个事件的概率. 也即, 在许多问

题中, 直接计算某个事件的概率是非常困难的, 但是一旦知道第二个事件发生与否, 就容易计算了. 我们接下来用一些例子阐述这点.

例 3a (第 1 部分) 保险公司认为人可以分为两类, 一类为容易出事故者, 另一类则为安全者. 他们的统计表明, 一个易出事故者在一年内发生事故的概率为 0.4, 而安全者, 这个概率则减少为 0.2, 若假定第一类人占人口的比例为 30%, 现有一个新的投保人来投保, 问该人在购买保单后一年内将出事故的概率有多大?

解: 以这个投保客户是不是易出事故的人作为条件, 我们将得到所求概率. 记 A_1 表示"投保客户一年内将出事故"这一事件, 而以 A 表示"投保人为容易出事故者"这一事件, 则所求概率 $P(A_1)$ 为

$$P(A_1) = P(A_1|A)P(A) + P(A_1|A^c)P(A^c) = 0.4 \times 0.3 + 0.2 \times 0.7 = 0.26 \quad \blacksquare$$

例 3a (第 2 部分) 假设一个新的投保人在购买保单后一年内出了事故, 问他是容易发生事故者的概率是多大?

解: 所求概率为 $P(A|A_1)$, 可从下式计算得到:

$$P(A|A_1) = \frac{P(AA_1)}{P(A_1)} = \frac{P(A)P(A_1|A)}{P(A_1)} = \frac{0.3 \times 0.4}{0.26} = 6/13 \quad \blacksquare$$

例 3b 考虑一副 52 张扑克牌的如下玩法: 将洗好的一副牌都扣住, 一次翻开一张. 玩家只有一次机会可以猜接下来翻开的一张是否是黑桃"A", 如果是, 那么玩家获胜; 如果不是, 那么玩家输. 另外, 如果一直到剩下一张还没有翻开, 而此前没有出现过黑桃"A", 且玩家也没有猜过, 那么玩家也获胜. 较好的策略? 较差的策略?

解: 其答案是: 任何一种策略, 获胜的概率都是 1/52. 为了说明这点, 我们将用归纳的方法证明这个结论: 对于 n 张牌, 其中有一张牌为黑桃"A", 那么不管采取何种策略, 获胜的概率都是 $1/n$. 这点显然对 $n = 1$ 是正确的. 假设对 $n-1$ 张牌, 该结论也成立. 现在考虑 n 张牌, 对于任一给定策略, 令 p 表示按该策略第一次就猜牌的概率. 如果第一次就猜牌, 那么获胜的概率为 $1/n$. 另一方面, 如果按策略第一次不猜牌, 那么获胜的概率就是第一张牌不是黑桃"A"的概率 $(n-1)/n$, 乘以在第一张牌不是黑桃"A"的条件下, 获胜的条件概率. 而此条件概率就等于含一张黑桃"A"的 $n-1$ 张牌的玩牌游戏中获胜的概率, 利用归纳假设, 该条件概率为 $1/(n-1)$, 因此, 按策略第一次不猜牌的条件下, 获胜的概率为 $\frac{n-1}{n}\frac{1}{n-1} = \frac{1}{n}$. 因此, 令 G 表示"第一次就猜牌"这一事件, 我们可得

$$P\{获胜\} = P\{获胜|G\}P(G) + P\{获胜|G^c\}(1 - P(G)) = \frac{1}{n}p + \frac{1}{n}(1-p) = \frac{1}{n} \quad \blacksquare$$

例 3c 在回答一道多项选择题时, 学生可能知道正确答案, 否则就猜一个. 令 p 表示他知道正确答案的概率, 则 $1-p$ 表示猜的概率. 假定学生猜中正确答案的

概率为 $1/m$, 此处 m 就是多项选择题的可选择答案数. 求在已知他回答正确的条件下, 该学生知道正确答案的概率?

解: 令 C 和 K 分别表示事件"该学生回答正确"和"该学生知道正确答案". 这样

$$P(K|C)=\frac{P(KC)}{P(C)}=\frac{P(C|K)P(K)}{P(C|K)P(K)+P(C|K^c)P(K^c)}$$
$$=\frac{p}{p+(1/m)(1-p)}=\frac{mp}{1+(m-1)p}$$

比如, $m=5, p=1/2$, 那么在已知该学生回答正确的条件下, 他知道正确答案的条件概率为 5/6. ■

例 3d 一项血液化验有 95% 的把握将患有某种疾病的患者诊断出来, 但是, 这项化验用于健康人也会有 1% 的"伪阳性"结果 (也即, 如果一个健康人接受这项化验, 则化验结果误诊此人患该疾病的概率为 0.01). 如果该疾病的患者事实上仅占人口的 0.5%, 若某人化验结果为阳性, 问此人确实患该疾病的概率为多大?

解: 以 D 表示"接受化验的这个人患该疾病"这一事件, E 表示"其化验结果为阳性"这一事件, 所求概率 $P(D|E)$ 为

$$P(D|E)=\frac{P(DE)}{P(E)}=\frac{P(E|D)P(D)}{P(E|D)P(D)+P(E|D^c)P(D^c)}$$
$$=\frac{0.95\times 0.005}{0.95\times 0.005+0.01\times 0.995}=\frac{95}{294}\approx 0.323$$

因此, 在验血结果为阳性的人当中, 真正患该病只有 32%. 对于这一结果, 许多学生感到非常吃惊 (因为验血似乎是个好办法, 他们总认为这个数值应该高得多), 因此, 有必要给出第二个解法. 与前一个解法比较, 第二个解法尽管不严格, 但却更直观.

由于事实上患该疾病的人占的人口比例为 0.5%, 平均地算, 接受化验的每 200 个人中应有 1 个患者, 而这项化验只能保证疾病的患者被诊断为患病的概率为 0.95, 因此, 平均地说, 每 200 个接受化验者能保证有 0.95 个人被诊断出, 并且此人真的患病. 但另一方面 (平均地说), 在其余 199 个健康人中, 这项化验会错误地诊断出 199×0.01 个人患该病, 因此, 每当诊断出 0.95 个病人时 (平均地说) 总有 199×0.01 个健康人误诊为患病. 于是, 当验血结果确定某人患该病时, 正确诊断所占比例为

$$\frac{0.95}{0.95+199\times 0.01}=\frac{95}{294}\approx 0.323$$ ■

公式 (3.1) 就是著名的全概公式. 利用它还可以导出著名的贝叶斯公式 (或逆概公式), 根据附加信息, 可对某事件的概率进行修正. 下面的例子就是贝叶斯公式的应用.

例 3e 假设某药剂师考虑如下诊断方案: 如果我有 80% 的可能确定病人确实有此病, 那么我会建议手术; 而如果我并不确定, 那么我会推荐做进一步的检查, 该

检查是昂贵的, 有时也是痛苦的. 现在, 开始我仅仅有 60% 的把握认为琼斯患有此病, 因此我推荐做了 A 项检查, 该检查对于确有此病的患者给出阳性结果, 而对健康人却不会给出阳性结果. 经检查琼斯的结果是阳性后, 正当我建议手术时, 琼斯给了我另一个信息, 他患有糖尿病. 这个信息带来麻烦. 尽管它并不影响我一开始认为他患有此病的 60% 的把握, 但是却影响了检查项目 A 的效果. 因为虽然该检查项目对健康人不给出阳性, 但是对于患有糖尿病却不患有这种疾病的人来说, 有 30% 的可能给出阳性结果. 那么我现在该如何做? 是做进一步检查, 还是立即手术?

解: 为了决定是否建议手术, 医生首先要计算在检查项目 A 为阳性结果的情况下, 琼斯患该病的概率. 令 D 表示"琼斯患此病"这一事件, E 表示"项目 A 为阳性结果"这一事件, 那么所求条件概率 $P(D|E)$ 为:

$$P(D|E)=\frac{P(DE)}{P(E)}=\frac{P(E|D)P(D)}{P(E|D)P(D)+P(E|D^c)P(D^c)}=\frac{0.6\times 1}{1\times 0.6+0.3\times 0.4}\approx 0.833$$

注意到我们以琼斯是否患有此病为条件计算了项目 A 为阳性结果的概率, 并且利用了如下事实: 因为琼斯患有糖尿病, 已知其不患上述疾病的条件下项目 A 为阳性结果的条件概率 $P(E|D^c)$ 等于 0.3, 因此, 医生现在能 80% 的把握确定琼斯患有此病, 所以应该建议手术. ■

例 3f 在某刑事调查过程中, 调查员有 60% 的把握认为嫌疑人确犯有此罪. 假定现在得到了一份新的证据, 表明罪犯有某个身体特征 (左撇子, 光头或者棕色头发等), 如果有 20% 的人有这种特征, 那么在嫌疑犯具有这种特征的条件下, 检查官认为他确犯此罪的把握为多大?

解: 令 G 表示"嫌疑犯的确犯此罪"这一事件, 而 C 表示"他具有罪犯的该身体特征", 那么我们有

$$P(G|C)=\frac{P(GC)}{P(C)}=\frac{P(C|G)P(G)}{P(C|G)P(G)+P(C|G^c)P(G^c)}=\frac{1\times 0.6}{1\times 0.6+0.2\times 0.4}\approx 0.882$$

其中我们假定了嫌疑犯事实上没犯罪却有该身体特征的概率为 0.2, 也就是具有这个特征的人口的比例. ■

例 3g 1965 年 5 月, 在布宜诺斯艾利斯举行的世界桥牌锦标赛上, 英国一对著名的桥牌手 T. 里斯和 B. 夏皮罗被指控作弊, 说是他们用手指作暗号暗示他们红心的张数. 里斯和夏皮罗都否认这项指控. 事后, 英国桥牌协会举行了一个听证会. 听证会以法律的程序进行, 既包含控方, 也包含辩方. 双方都有目击证人. 在接下来的调查过程中, 控方检查了里斯和夏皮罗打的几手牌, 并声称他们的打法与用作弊已知了红心的张数的打法是吻合的. 针对这个观点, 辩方律师指出, 他们的打法同样也同标准打法一致. 然而, 控方指出, 既然他们的打法与其作弊的假设是一致的, 那么就应该是支持这种假设. 你如何理解控方的理由?

解: 此问题基本上是新的证据 (在此例中, 牌的打法) 是如何影响某个特定假

设成立的概率的问题. 现在, 令 H 表示"某个特定假设 (里斯和夏皮罗确实作弊)", 而 E 表示"新的证据", 那么

$$P(H|E) = \frac{P(HE)}{P(E)} = \frac{P(E|H)P(H)}{P(E|H)P(H) + P(E|H^c)[1-P(H)]} \tag{3.2}$$

其中 $P(H)$ 为在新证据展示之前我们对假设成立的可能性的估值. 新证据支持假设成立, 如果它使得假设成立的可能性增大, 也即 $P(H|E) \geqslant P(H)$. 利用公式 (3.2), 此式等价于 $P(E|H) \geqslant P(E|H)P(H) + P(E|H^c)[1-P(H)]$, 或者等价于 $P(E|H) \geqslant P(E|H^c)$. 也就是说, 任何新的证据被认为是支持假设成立的, 除非当假设成立时, 该证据发生的概率大于假设不成立时发生的概率. 事实上, 已知新的证据发生的条件下, 假设成立的新概率和初始概率的关系可从下式看出:

$$P(H|E) = \frac{P(H)}{P(H) + [1-P(H)]\dfrac{P(E|H^c)}{P(E|H)}}$$

因此, 在所考虑的问题中, 牌手的打法可以认为是支持假设成立的, 除非在他们作弊的条件下这种打法的可能性大于他们不作弊的条件下这种打法的可能性. 而控方并没有作此声明. 因此, 他们关于新证据是支持作弊的假设的这一断言是无效的. ■

在一咖啡店喝冰茶时, 我要了一杯水和同样杯子的一杯茶. 喝茶时, 我不断地向茶杯里续水. 假设水和茶充分混合, 那么关于我的最后一口是茶的概率问题引出了以下问题中的 (a), 并且给出了有趣的答案.

例 3h 坛子 1 里面最初有 n 个红色分子, 坛子 2 里面有 n 个蓝色分子, 按照以下方式进行操作: 从坛子 1 里随机移走一个分子, 然后, 从坛子 2 里取一个分子 (如果里面还有分子的话) 放进坛子 1 里. 一直进行这样的操作, 直到坛子 1 和坛子 2 中所有的分子都被移走 (一共从坛子 1 移了 $2n$ 次, 但是从坛子 2 一共移走了 n 个蓝分子.).

(a) 求 $P(R)$, 其中 R 表示事件"从坛子 1 最后一个移走的分子是红色的".

(b) 如果坛子 1 里最初有 r_1 个红分子, b_1 个蓝分子, 而坛子 2 里有 r_2 个红分子, b_2 个蓝分子, 重求上述概率.

解: (a) 将注意力放在某个特殊的红分子上, 令 F 表示"该特殊的红分子是最后一个被移走的". 若 F 发生, 那么在坛子 1 中移走了 n 个分子以后 (相应坛子 2 中的 n 个蓝色分子也已经被移走了), 这个分子仍然在坛子 1 中. 令 N_i 表示"该分子不是第 i 次被移走的分子", 显然有

$$\begin{aligned}P(F) &= P(N_1\cdots N_nF) = P(N_1)P(N_2|N_1)\cdots P(N_n|N_1\cdots N_{n-1})P(F|N_1\cdots N_n)\\ &= \left(1-\frac{1}{n}\right)\cdots\left(1-\frac{1}{n}\right)\frac{1}{n}\end{aligned}$$

其中 $P(F|N_1,\cdots,N_n)$ 表示坛子 1 中共有 n 个分子 (包括那个特殊的红色分子),

一个一个取出时, 那个特殊分子在最后取出, 显然它等于 $1/n$.

因此, 如果我们给 n 个红分子标号, 令 R_j 表示红分子 j 最后被移走, 那么通过上面的分析可得

$$P(R_j) = \left(1 - \frac{1}{n}\right)^n \frac{1}{n}$$

因为事件 R_j 互不相容, 我们可得

$$P(R) = P\left(\sum_{j=1}^{n} R_j\right) = \sum_{j=1}^{n} P(R_j) = \left(1 - \frac{1}{n}\right)^n \approx \mathrm{e}^{-1}$$

(b) 现在假设坛子 i 里最初有 r_i 个红分子和 b_i 个蓝分子 $(i = 1, 2)$. 为了计算最后移走的分子是红分子的概率 $P(R)$, 将注意力集中在坛子 1 里最初的某个特殊的分子 (这个特殊的分子可以是红的, 也可以是蓝的). 类似 (a) 可得该分子在最后被移走的概率为

$$p = \left(1 - \frac{1}{r_1 + b_1}\right)^{r_2 + b_2} \frac{1}{r_1 + b_1}$$

上式中, 因子 $\left(1 - \dfrac{1}{r_1 + b_1}\right)^{r_2 + b_2}$ 表示当第二个坛子被取空以后, 那个特殊的分子仍然在第一个坛子内的概率, 而 $1/(r_1 + b_1)$ 为在前面那个事件发生的条件下, 继续从坛子 1 内一个一个地取分子, 而那个特殊的分子被最后取出的概率. 现在记 O 为 "最后移走的分子是在坛子 1 中的分子", 则

$$P(O) = (r_1 + b_1)p = \left(1 - \frac{1}{r_1 + b_1}\right)^{r_2 + b_2}$$

为计算 $P(R)$, 我们以 O 是否发生为条件, 得到

$$\begin{aligned} P(R) &= P(R|O)P(O) + P(R|O^{\mathrm{c}})P(O^{\mathrm{c}}) \\ &= \frac{r_1}{r_1 + b_1}\left(1 - \frac{1}{r_1 + b_1}\right)^{r_2 + b_2} + \frac{r_2}{r_2 + b_2}\left[1 - \left(1 - \frac{1}{r_1 + b_1}\right)^{r_2 + b_2}\right] \end{aligned}$$

如果 $r_1 + b_1 = r_2 + b_2 = n$, 这样两个坛子里最初都有 n 个分子, 当 n 充分大时,

$$P(R) \approx \frac{r_1}{r_1 + b_1}\mathrm{e}^{-1} + \frac{r_2}{r_2 + b_2}(1 - \mathrm{e}^{-1})$$ ■

当发现新的证据时, 假设成立的概率之变化可以表示为假设的"优势"之变化, 其中优势的概念定义如下.

定义 事件 A 的优势定义为

$$\frac{P(A)}{P(A^{\mathrm{c}})} = \frac{P(A)}{1 - P(A)}$$

即事件 A 的优势告诉我们该事件发生的可能性是不发生的可能性的倍数. 举例来说, 如果 $P(A) = 2/3$, 那么 $P(A) = 2P(A^{\mathrm{c}})$, 因此, 事件 A 的优势等于 2. 如果某事件的优势等于 α, 那么通常称支持假设成立的优势为 "α 比 1" 也是同样的意思.

现在考虑假设 H 以概率 $P(H)$ 成立, 如果我们发现了新的证据 E, 那么 E 成立的条件下, H 成立和 H 不成立的条件概率分别为

$$P(H|E)=\frac{P(E|H)P(H)}{P(E)}\qquad P(H^c|E)=\frac{P(E|H^c)P(H^c)}{P(E)}$$

因此, 引进证据 E 后, 假设 H 的优势为

$$\boxed{\frac{P(H|E)}{P(H^c|E)}=\frac{P(H)}{P(H^c)}\frac{P(E|H)}{P(E|H^c)}}\tag{3.3}$$

即 H 的新的优势值等于它原来的优势值乘以新的证据在 H 和 H^c 之下的条件概率比值. 这个结论也验证了例 3f 的结论. 如果, H 的条件下新的证据的条件概率大于 H^c 的条件下的条件概率时, H 的优势值是递增的. 反之, 为递减的. 此处我们将随机事件 H 称为"假设", 这是借用了犯罪学中的术语. 我们称犯罪嫌疑人"有罪"为假设, 而不称事件.

例 3i 坛子中有硬币 A 和硬币 B 各一枚. 当抛掷硬币 A 时, 其正面朝上的概率为 1/4, 而抛掷硬币 B 时, 正面朝上的概率为 3/4. 假设从坛子中随机挑选一枚硬币进行抛掷. 如果已知抛掷的结果是正面朝上, 那么选中的是硬币 A 的概率是多大?

解: 记 A 为事件"抛掷的硬币为 A", $B=A^c$ 表示"抛掷的硬币为 B". 我们需要计算 $P(A|\text{正面向上})$. 利用公式 (3.3) 得到

$$\begin{aligned}\frac{P(A|\text{正面向上})}{P(B|\text{正面向上})}&=\frac{P(A)}{P(B)}\frac{P(\text{正面向上}\mid A)}{P(\text{正面向上}\mid B)}\\&=\frac{2/3}{1/3}\cdot\frac{1/4}{3/4}=2/3\end{aligned}$$

即在正面向上的条件下, 硬币 A 的优势为 2/3:1, 或, 抛掷硬币 A 的 (条件) 概率为 2/5. ■

公式 (3.1) 可以推广如下: 假设 $F_1,F_2,\cdots,F_n$ 为互不相容事件, 满足 $\bigcup_{i=1}^n F_i=S$, 也就是说, $F_1,F_2,\cdots,F_n$ 中必有且仅有一个事件发生. 将 E 写成 $E=\bigcup_{i=1}^n EF_i$. 然后利用事实: $EF_i,i=1,\cdots,n$ 为互不相容事件, 我们可得

$$P(E)=\sum_{i=1}^n P(EF_i)=\sum_{i=1}^n P(E|F_i)P(F_i)\tag{3.4}$$

公式 (3.4) 表明, 对于给定的一组完备的事件 $F_1,F_2,\cdots,F_n$, 即 $F_1,\cdots,F_n$ 为互不相容的事件, 并且 $\bigcup F_i=S$(S 为必然事件). 我们可以通过 $P(E|F_i)$ 来计算 $P(E)$. 即公式 (3.4) 叙述了 $P(E)$ 等于 $P(E|F_i)$ 的加权平均, 每项的权为事件 F_i 发生的概率.

例 3j 在第 2 章例 5j 中, 讨论了一事件的概率计算问题. 设将一副扑克牌一张一张地往外翻, 在翻出第一张 A 之后接着再翻出某一种牌的概率的计算. 我们利

用组合方法指出这个概率等于 1/52. 现在利用条件概率的方法进行计算. 令 E 为事件“翻出第一张 A 后又接着翻出某一牌 (例如 x)”. 将牌 x 去掉, 记 O 为剩下的 51 张牌的随机次序, 我们有

$$P(E)=\sum_{\mathrm{O}} P(E|\mathrm{O})P(\mathrm{O})$$

对于给定的 O, 对应于 52 个整副牌的次序, 它们相应于把 x 插入由 O 所形成的 52 个位置. 显然这 52 个位置是完全等概率的, 而事件 E 只相应于 x 插在第一张 A 后面的位置, 这样 $P(E|\mathrm{O})=1/52$. 这说明 $P(E)=1/52$. ■

现在假定, $F_1,\cdots F_n$ 是一组互不相容的事件, 并且它们的和事件为必然事件 (称为完备事件组).

现在假设 E 发生了 (新的证据), 我们想要计算 F_j 发生的概率, 利用公式 (3.4), 我们有如下命题.

命题 3.1

$$P(F_j|E)=\frac{P(EF_j)}{P(E)}=\frac{P(E|F_j)P(F_j)}{\sum_{i=1}^{n} P(E|F_i)P(F_i)} \tag{3.5}$$

公式 (3.5) 称为贝叶斯公式, 最早由英国哲学家托马斯 · 贝叶斯提出. 如果我们把事件 F_j 设想为关于某件事件的各个可能的“假设条件”, 那么, 贝叶斯公式可以这样理解: 它告诉我们, 在试验之前对这些假设条件所作的判断 [即 $P(F_j)$], 可以根据试验的结果来进行修正.

例 3k 一架飞机失踪了, 推测它等可能地坠落在 3 个区域. 令 $1-\beta_i$ 表示飞机坠落在第 i 个区域时被发现的概率 (β_i 称为疏忽概率, 它决定于该区域的地理和环境条件). 已知对区域 1 的搜索没有发现飞机, 求在此条件下, 飞机坠落在第 i 个区域 $(i=1,2,3)$ 的条件概率.

解: 令 $R_i, i=1,2,3$ 表示“飞机坠落在第 i 个区域”这一事件, 令 E 表示“对第 1 个区域的搜索没有发现飞机”这一事件, 利用贝叶斯公式可得

$$P(R_1|E)=\frac{P(ER_1)}{P(E)}=\frac{P(E|R_1)P(R_1)}{\sum_{i=1}^{3} P(E|R_i)P(R_i)}=\frac{\beta_1\times\frac{1}{3}}{\beta_1\times\frac{1}{3}+1\times\frac{1}{3}+1\times\frac{1}{3}}=\frac{\beta_1}{\beta_1+2}$$

对于 $j=2,3$, 有

$$P(R_j|E)=\frac{P(E|R_j)P(R_j)}{P(E)}=\frac{1\times\frac{1}{3}}{\beta_1\times\frac{1}{3}+\frac{1}{3}+\frac{1}{3}}=\frac{1}{\beta_1+2}\qquad j=2,3$$

值得指出的是，当搜索了第 1 个区域没有发现飞机时，飞机坠落在第 $j(j\neq 1)$ 的概率会增大，而坠落在第 1 个区域的概率会减小. 这是一个常识问题：因为既然在第 1 个区域没有发现飞机，当然飞机坠落在该区域的概率会减少，而坠落在其他区域的概率会增大. 而且飞机坠落在第 1 个区域的条件概率是疏忽概率 β_1 的递增函数. 当 β_1 增加时，增大了飞机坠落在第 1 个区域的条件概率. 类似地，$P(R_j|E), j\neq 1$ 是 β_1 的递减函数. ■

下一个例子经常被学过概率而又不讲道德的学生们用来从他们概率知识较少的朋友那里赢钱.

例 3l 假设有 3 张形状完全相同但所涂颜色不同的卡片，第一张两面全是红色，第二张两面全是黑色，而第三张是一面红一面黑. 将这 3 张卡片放在帽子里混合后，随机地取出 1 张放在地上，如果取出的卡片朝上的一面是红色的，那么另一面为黑色的概率是多大?

解：令 RR, BB, RB 分别表示取出的卡片是"两面红"、"两面黑"以及"一面红一面黑"这三个事件. 再令 R 表示取出的卡片"朝上一面是红色"这一事件. 我们可如下得到所求概率：

$$
\begin{aligned}
P(RB|R) &= \frac{P(RB\cap R)}{P(R)} \\
&= \frac{P(R|RB)P(RB)}{P(R|RR)P(RR)+P(R|RB)P(RB)+P(R|BB)P(BB)} \\
&= \frac{\frac{1}{2}\times\frac{1}{3}}{1\times\frac{1}{3}+\frac{1}{2}\times\frac{1}{3}+0\times\frac{1}{3}} = \frac{1}{3}
\end{aligned}
$$

因此，反面是黑色的概率为 1/3. 但是有些学生猜另一面是黑色的概率为 1/2. 他们的错误在于认为这张牌有两种可能，两面全红，或者一面红一面黑，这两种可能性是一样的. 事实上，你可以把三张牌的 6 个面记为 $(R_1,R_2),(B_1,B_2),(R_3,B_3)$，其中 (R_1,R_2) 表示全红的那张牌的两个面，(B_1,B_2) 表示两面均为黑色的那张牌的两个面，(R_3,B_3) 表示一红一黑的那张牌的两个面. 即 6 个面是以相等的概率出现的. 只有 R_3 出现时，背面是黑色的. 而 R_1,R_2 出现时，背面均为红色的. 因此，当正面为红色时，背面为黑色的概率为 1/3 而不是 1/2. ■

例 3m 镇上新搬来一对夫妇，已知他们有两个孩子. 假设某天遇到该母亲带着一个女儿在散步，问她的两个孩子都是女儿的概率是多大?

解：首先，定义如下事件.

G_1：第一个 (也即最大的) 孩子为女孩;

G_2：第二个孩子为女孩;

G：被看到跟母亲一起散步的为女孩.

而且令 B_1, B_2, B 表示类似上述的事件, 其中"女孩"替换为"男孩". 这样, 所求概率 $P(G_1G_2|G)$ 可以表示如下:

$$P(G_1G_2|G) = \frac{P(G_1G_2G)}{P(G)} = \frac{P(G_1G_2)}{P(G)}$$

而且,

$$\begin{aligned}P(G) =& P(G|G_1G_2)P(G_1G_2) + P(G|G_1B_2)P(G_1B_2) \\ &+ P(G|B_1G_2)P(B_1G_2) + P(G|B_1B_2)P(B_1B_2) \\ =& P(G_1G_2) + P(G|G_1B_2)P(G_1B_2) + P(G|B_1G_2)P(B_1G_2)\end{aligned}$$

其中最后一个等式用到了结论 $P(G|G_1G_2) = 1$ 以及 $P(G|B_1B_2) = 0$. 如果我们通常地假设 4 种可能结果 G_1, G_2, B_1, B_2 都是等可能的, 那么我们有

$$P(G_1G_2|G) = \frac{1/4}{1/4 + P(G|G_1B_2)/4 + P(G|B_1G_2)/4} = \frac{1}{1 + P(G|G_1B_2) + P(G|B_1G_2)}$$

因此, 答案依赖于已知事件 G_1B_2 条件下, 碰到母亲带着女孩的条件概率, 以及已知事件 G_2B_1 条件下, 碰到和母亲带着男孩的条件概率. 举例来说, 我们假定母亲对于性别没有倾向, 母亲带着大孩子一起散步的概率为 p, 那么有

$$P(G|G_1B_2) = p = 1 - P(G|B_1G_2)$$

即 $P(G|G_1B_2) + P(G|B_1G_2) = 1$, 从而 $P(G_1G_2|G) = 1/2$. 另一面, 如果假定两个孩子性别不同, 母亲带着女孩一起散步的概率为 q, 且与孩子出生的次序是独立的, 那么我们有 $P(G|G_1B_2) = P(G|B_1G_2) = q$, 意味着 $P(G_1G_2|G) = 1/(1+2q)$. 举例来说, 如果我们取 $q = 1$, 即母亲总是选择女孩一起去散步, 那么两个都是女孩的条件概率为 1/3. 这点同 例 2b 是一致的, 因为事件"看见母亲带着一个女孩"与事件"至少有一个女孩"是等价的.

因此, 综上所述, 此问题是无法解的. 事实上, 即使假设孩子的性别是等可能的, 我们仍需要额外的假设才能解决问题. 因为该试验的样本空间包含了如下形式的向量: (s_1, s_2, i), 其中 s_1 表示大孩子的性别, s_2 表示小孩子的性别, i 表示被碰到的孩子的出生次序. 因此, 为确定样本空间的点的概率, 光知道孩子的性别的概率是不够的, 还需要知道母亲所带孩子的出生次序的条件概率 (给定大小孩子的性别). ■

例 3n 储物箱里有 3 种不同的一次性手电. 第一种手电使用超过 100 小时的概率为 0.7, 而第二种和第三种手电相应的概率分别只有 0.4 和 0.3. 假设箱子里的手电, 20% 的为第一种, 第二种和第三种分别为 30% 和 50%.

(a) 随机挑一个手电, 能使用 100 小时以上的概率是多大?

(b) 已知手电使用超过 100 小时, 问它是第 j 种手电的条件概率是多大 ($j = 1, 2, 3$)?

解: (a) 令 A 表示"挑出的手电能使用 100 小时"这一事件, 令 F_j 表示挑出了第 j 种手电 $(j=1,2,3)$. 为了计算 $P(A)$, 以挑出手电的种类为条件, 可得

$$P(A)=P(A|F_1)P(F_1)+P(A|F_2)P(F_2)+P(A|F_3)P(F_3)$$
$$=0.7\times0.2+0.4\times0.3+0.3\times0.5=0.41$$

因此, 随机挑选的手电能使用使用 100 小时以上的概率为 0.41.

(b) 利用贝叶斯公式得到所求概率:

$$P(F_j|A)=\frac{P(AF_j)}{P(A)}=\frac{P(A|F_j)P(F_j)}{0.41}$$

因此,

$$P(F_1|A)=0.7\times0.2/0.41=14/41$$
$$P(F_2|A)=0.4\times0.3/0.41=12/41$$
$$P(F_3|A)=0.3\times0.5/0.41=15/41$$

举例来说, 对第一种手电, 尽管被选中的初始概率只有 0.2, 但是得到手电使用超过 100 小时的信息后, 这个概率提升到 $14/41\approx0.341$. ■

例 3o 警方从案发现场罪犯的身体遗留物中提取了 DNA, 法医研究后注意到能够辨认的只有 5 对染色体, 而且每个无罪的人, 与这 5 对染色体相匹配的概率为 10^{-5}, 律师认为罪犯就是该城镇 100 万居民之一. 在过去 10 年内, 该城镇有 10 000 人曾蹲过监狱. 他们的 DNA 资料都记录在案. 在检查这些 DNA 文档之前, 律师认为这 10 000 有犯罪前科的人犯此罪的概率为 α. 而其余 990 000 居民, 每人犯此罪的概率为 β, 其中 $\alpha=c\beta$(也即, 他认为最近 10 年内释放的有犯罪前科的人作案的可能性是其他人的 c 倍). 将 DNA 分析结果同这 10 000 个有犯罪前科的人的数据文档对比后, 发现只有 AJ 琼斯的 DNA 符合. 假设律师关于 α 和 β 之间的关系是准确的, AJ 作案的可能性有多大?

解: 首先注意到概率之和必等于 1, 因此我们有

$$1=10\ 000\alpha+990\ 000\beta=(10\ 000c+990\ 000)\beta$$

即

$$\beta=\frac{1}{10\ 000c+990\ 000}\qquad\alpha=\frac{c}{10\ 000c+990\ 000}$$

现在, 令 G 表示"AJ 为作案者", 令 M 表示"AJ 是这 10 000 人中唯一的与现场 DNA 相匹配的人". 那么

$$P(G|M)=\frac{P(GM)}{P(M)}=\frac{P(G)P(M|G)}{P(M|G)P(G)+P(M|G^c)P(G^c)}$$

如果 AJ 为作案者, 此时其他的 9999 人都不是作案者. 这 10 000 人中 AJ 是唯一匹配者的概率为 $P(M|G)=(1-10^{-5})^{9999}$.

令 $C=\{$除 AJ 外, 其他有前科的人不是作案者$\}$, 则

$$P(C|G^c)=\frac{P(CG^c)}{P(G^c)}=\frac{1-10\ 000\alpha}{1-\alpha}$$

同样, 在除 AJ 外, 其他有前科的人不是作案者的前提下, 这 9999 人与现场 DNA 相匹配的概率为 $(1-10^{-5})^{9999}$, 所以[①]

$$P(M|G^c)=10^{-5}(1-10^{-5})^{9999}\left(\frac{1-10\ 000\alpha}{1-\alpha}\right)$$

现在我们已经计算得到 $P(M|G)$ 和 $P(M|G^c)$ 的公式, 结合 $P(G)=\alpha$, 将这些公式代入到 $P(G|M)$ 的表达式中, 得到

$$P(G|M)=\frac{\alpha}{\alpha+10^{-5}(1-10\ 000\alpha)}=\frac{1}{0.9+\dfrac{10^{-5}}{\alpha}}$$

因此, 如果律师最初认为任一有犯罪前科的人作案的可能性是没有犯罪前科的人的 100 倍 (也即 $c=100$), 那么 $\alpha=1/19\ 900$, 且 $P(G|M)=1/1.099\approx 0.9099$. 如果律师最初认为 $c=10$, 那么 $\alpha=1/109\ 000$, 且 $P(G|M)=1/1.99\approx 0.5025$. 如果律师最初认为任一有犯罪前科的人作案的可能性与镇里其他人是相同的 $(c=1)$, 那么 $\alpha=10^{-6}$, 且 $P(G|M)=1/10.9\approx 0.0917$. 因此, 概率变化范围是从大概 9%(此时律师最初假设所有人作案的概率一样) 到 91%(此时他认为每个有犯罪前科的人作案的概率为其他任一居民的 100 倍). ■

3.4 独立事件

本章前面的例子显示: 已知 F 的条件下 E 发生的条件概率 $P(E|F)$ 一般来说不等于 E 发生的 (无条件) 概率 $P(E)$. 也就是说, 知道了 F 已发生通常会改变 E 的发生机会. 但在一些特殊情形下, $P(E|F)$ 确实等于 $P(E)$, 此时我们称 E 和 F

---

① 文中的公式可以这样理解: 对于概率 $P(M|G^c)$ 利用全概公式可得

$$P(M|G^c)=P(C|G^c)P(M|CG^c)+P(C^c|G^c)P(M|C^cG^c)$$

上述右边第二项中因子 $P(M|C^cG^c)$ 是不可能事件的概率, 这是因为 C^cG^c 表示作案者在除 AJ 外的 9 999 个有前科的人中间, 而 M 表示这 10 000 个人中 AJ 是唯一匹配者, 故 $P(M|C^cG^c)=0$. 这样全概公式转化为

$$P(M|G^c)=P(C|G^c)P(M|CG^c).$$

而 CG^c 表示这 10 000 个有前科者均不是作案者, 这样 AJ 为唯一匹配者的概率为 $P(M|CG^c)=10^{-5}(1-10^{-5})^{9999}$. 而正文中前面已求得 $P(C|G^c)$ 的值. 将这两个因子相乘就可得到文中的公式, 即 $P(M|G^c)$ 的值. —— 译者注

独立. 即如果已知 F 的发生并不影响 E 发生的可能性, 那么 E 和 F 就是独立的.

因为 $P(E|F)=P(EF)/P(F)$, 所以如果下式成立, 那么 E 和 F 独立:

$$P(EF) = P(E)P(F) \tag{4.1}$$

由于公式 (4.1) 关于 E 和 F 是对称的, 就说明了只要 E 和 F 独立, 那么 F 和 E 也独立, 因此, 我们有以下定义.

定义 对于两个事件 E 和 F, 若 (4.1) 式成立, 则称它们是**独立的** (independent. 若两个事件 E 和 F 不独立, 则称它们是**相依的**(dependent), 或相互不独立.

例 4a 从一副洗好的 52 张扑克牌里随机抽取一张牌. 记 E 表示事件"抽取的牌为一张'A'", 记 F 表示事件"抽取的牌为一张黑桃", 那么 E 和 F 就是独立的. 因为 $P(EF) = 1/52$, 而 $P(E) = 4/52$ 且 $P(F) = 13/52$. ■

例 4b 掷两枚硬币, 假设全部 4 个结果出现的可能性是一样的. 记 E 表示事件"第一枚硬币正面朝上", 记 F 表示"第二枚硬币反面朝上", 那么 E 和 F 是独立的, 因为 $P(EF) = P\{(\mathrm{H},\mathrm{T})\} = 1/4$, 而 $P(E) = P(\{(\mathrm{H},\mathrm{H}),(\mathrm{H},\mathrm{T})\}) = 1/2$, 且 $P(F) = P(\{(\mathrm{H},\mathrm{T}),(\mathrm{T},\mathrm{T})\}) = 1/2$. ■

例 4c 掷两枚均匀的骰子, 记 E_1 表示事件"骰子点数之和为 6", F 表示事件"第一枚骰子点数为 4", 那么 $P(E_1F) = P(\{(4,2)\}) = 1/36$. 而 $P(E_1)P(F) = 5/36 \times 1/6 = 5/216$, 因此, E_1 和 F 不独立. 直观来说, 这个原因是显而易见的, 因为如果我们关注掷出点数和为 6(两枚骰子), 那么在第一枚骰子为 4(或者 1,2,3,4,5 中任一个), 我们都还乐观, 因为此时仍有机会得到和为 6. 另一方面, 如果第一枚骰子为 6, 那么我们就不乐观了, 因为没有任何机会得到点数和为 6 了. 也就是说, 得到和为 6 的机率依赖于第一枚骰子的结果, 因此, E_1 和 F 不可能独立.

现在, 令 E_2 表示事件"骰子数之和为 7", 那么 E_2 是否和 F 独立? 答案是肯定的, 因为 $P(E_2F) = P(\{(4,3)\}) = 1/36$, 而 $P(E_2)P(F) = \frac{1}{6} \times \frac{1}{6} = 1/36$.

我们留给读者去证明此直观的结论: 为什么事件"骰子点数之和为 7"同第一枚骰子的点数是独立的. ■

例 4d 令 E 表示事件"下届总统是共和党人", F 表示事件"未来一年将会有一次大地震", 大多数人会认为 E 和 F 是独立的. 然而, 如果另外一个事件 G 是"选举后两年之内会经历经济衰退", 那么对于 E 和 G 是否独立, 却存在着长期争论. ■

接下来证明如果 E 独立于 F, 那么 E 也独立于 F^c.

命题 4.1 如果 E 和 F 独立, 那么 E 和 F^c 也独立.

证明: 假定 E 和 F 独立. 由于 $E = EF \cup EF^c$, 且 EF 和 EF^c 显而易见

是互不相容的, 因此有 $P(E) = P(EF) + P(EF^c) = P(E)P(F) + P(EF^c)$ 这样, $P(EF^c) = P(E)[1 - P(F)] = P(E)P(F^c)$ 命题得以证明. □

因此, 如果 E 和 F 是独立的, 那么无论得知 F 发生的信息, 还是得知 F 不发生的信息, E 发生的概率都是不变的.

现在假设 E 既和 F 独立, 也和 G 独立, 那么 E 是否一定和 FG 独立? 答案有点不可思议, 是否定的. 考虑如下例子:

例 4e 掷两枚均匀的骰子. 记 E 表示事件"骰子点数之和为 7", 记 F 表示事件"第一枚骰子点数为 4"以及 G 表示事件"第二枚骰子点数为 3". 从例 4c 可以得知, E 和 F 是独立的, 同样的理由可以说明 E 和 G 也是独立的. 但是, 很明显 E 和 FG 是不独立的, 因为 $P(E|FG) = 1$. ■

从例 4e 还可以得出关于三个事件 E, F 和 G 的独立性的合适的定义. 三个事件的独立性比要求所有的 $\binom{3}{2}$ 对事件的独立更强, 我们给出如下定义.

定义 三个事件 E, F 和 G 称为独立的, 如果

$$P(EFG) = P(E)P(F)P(G) \qquad P(EF) = P(E)P(F)$$
$$P(EG) = P(E)P(G) \qquad P(FG) = P(F)P(G)$$

值得注意的是, 如果 E, F 和 G 是独立的, 那么 E 与 F 和 G 的任意组合事件都是独立的. 比如, E 和 $F \cup G$ 就是独立的, 因为

$$\begin{aligned}P[E(F \cup G)] &= P(EF \cup EG) = P(EF) + P(EG) - P(EFG) \\ &= P(E)P(F) + P(E)P(G) - P(E)P(FG) \\ &= P(E)[P(F) + P(G) - P(FG)] \\ &= P(E)P(F \cup G)\end{aligned}$$

当然, 还可以推广到三个以上事件的独立性定义. 事件 $E_1, \cdots, E_n$ 称为独立的, 如果对这些事件的任意子集 $E_{1'}, E_{2'}, \cdots, E_{r'}, r' \leqslant n$, 都有

$$P(E_{1'}E_{2'}\cdots E_{r'}) = P(E_{1'})P(E_{2'})\cdots P(E_{r'})$$

最后, 我们来定义无限个事件的独立性. 如果无限个事件的任意有限个子集都是独立的, 则称这无限个事件是相互独立的.

有时会遇到这种情况, 所考虑的概率试验由一系列子试验组成. 例如, 连续抛掷一枚硬币这个试验, 就可以把每掷一次看作一个子试验. 在许多场合下, 假定任一组子试验的结果不影响其他子试验的结果的假设是合理的. 如果真是这样, 我们称这些子试验是相互独立的. 更确切的说, 称一系列子试验是相互独立的, 如果任意的事件序列 $E_1, E_2, \cdots, E_n, \cdots$ 是相互独立的事件序列, 这里, 事件 E_i 完全由第

i 次子试验的结果所决定.

如果各个子试验彼此相同, 即各子试验有相同的 (子) 样本空间及相同的事件概率函数, 那么就称这些试验为**重复试验**.

例 4f 设在独立重复试验的无限序列中, 每次试验结果成功的概率为 p, 失败的概率为 $1-p$, 试求如下概率:

(a) 前 n 次试验中至少成功 1 次;

(b) 前 n 次试验中成功 k 次;

(c) 所有试验结果都成功.

解: 为计算前 n 次试验中至少成功 1 次的概率, 先求它的对立事件的概率. 它的对立事件就是"前 n 次试验全失败了". 若以 E_i 表示第 i 次试验失败这一事件, 则由独立性可得, 前 n 次试验全失败的概率

$$P(E_1E_2\cdots E_n)=P(E_1)P(E_2)\cdots P(E_n)=(1-p)^n$$

因此, (a) 的答案就是 $1-(1-p)^n$.

为解 (b), 考虑任一个由 k 个成功、$n-k$ 个失败组成的前 n 个结果的特定序列. 由独立性知, 每个这样的序列发生的概率为 $p^k(1-p)^{n-k}$. 由于共有 $\binom{n}{k}$ 个这样的序列 (由 k 个成功与 $n-k$ 个失败组成的序列总数为 $n!/[k!(n-k)!]$), 故所求概率为

$$P\{\text{恰有}k\text{次成功}\}=\binom{n}{k}p^k(1-p)^{n-k}$$

为解 (c), 我们由 (a) 注意到, 前 n 次试验全成功的概率为 $P(E_1^cE_2^c\cdots E_n^c)=p^n$. 因此, 运用概率的连续性属性 (2.6 节), 我们可得所求概率 $P\left(\bigcap_{i=1}^{\infty}E_i^c\right)$ 为

$$P\left(\bigcap_{i=1}^{\infty}E_i^c\right)=P\left(\lim_{n\to\infty}\bigcap_{i=1}^{n}E_i^c\right)=\lim_{n\to\infty}P\left(\bigcap_{i=1}^{n}E_i^c\right)$$

$$=\lim_n p^n=\begin{cases}0 & \text{如果}\quad p<1\\ 1 & \text{如果}\quad p=1\end{cases}$$ ■

例 4g 由 n 个元件组成的系统称为并联的, 如果至少有一个元件工作正常, 那么整个系统都工作正常 (如图 3.2). 对于这样的系统, 如果元件 i 工作正常的概率为 $p_i, i=1,\cdots,n$, 并且各元件的工作状态相互独立. 那么整个系统工作正常的概率是多大?

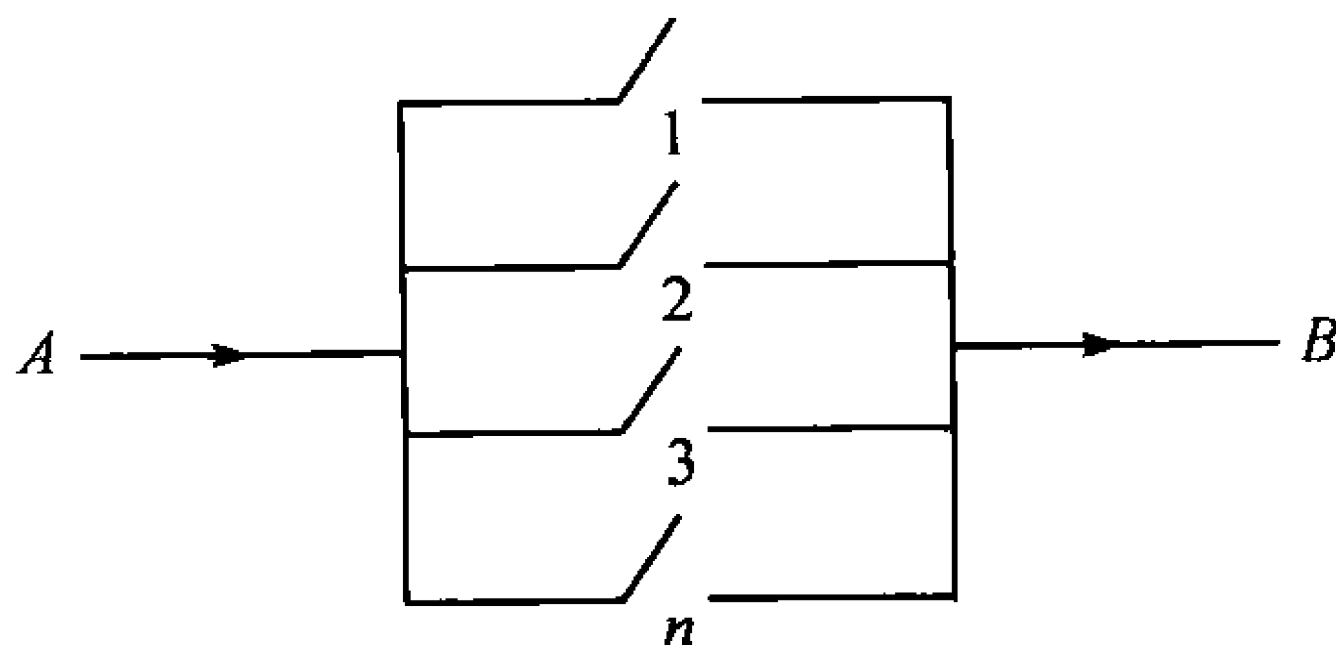


图3.2 并联系统: 只要有一个开关是通的, A与B之间就是通的

解: 令 A_i 表示事件“元件 i 工作正常”, 那么

$$\begin{aligned}P\{\text{系统工作正常}\} &= 1 - P\{\text{系统工作不正常}\} \\ &= 1 - P\{\text{所有元件工作不正常}\} \\ &= 1 - P\Big(\bigcap_i A_i^c\Big) = 1 - \prod_{i=1}^{n}(1-p_i) \qquad \text{利用独立性}\end{aligned}$$

■

例 4h 进行独立重复试验, 每次试验为掷两枚均匀的骰子, 并记录两枚骰子点数之和. 问“和为 5”出现在“和为 7”之前的概率是多少?

解: 令 E_n 表示事件“前 $n-1$ 次试验中, 5 和 7 都不出现, 而第 n 次试验出现 5”, 那么所求概率为

$$P\Big(\bigcup_{n=1}^{\infty} E_n\Big) = \sum_{n=1}^{\infty} P(E_n)$$

因为任一次试验中, $P(\text{和为}5) = 4/36$, 且 $P(\text{和为}7) = 6/36$, 这样, 利用试验的独立性可以得到

$$P(E_n) = \left(1 - \frac{10}{36}\right)^{n-1} \times \frac{4}{36}$$

因此,

$$P\left(\bigcup_{n=1}^{\infty} E_n\right) = \frac{1}{9}\sum_{n=1}^{\infty}\left(\frac{13}{18}\right)^{n-1} = \frac{1}{9} \times \frac{1}{1 - 13/18} = \frac{2}{5}$$

该结果还可以利用条件概率得到. 令 E 表示事件“和为 5 出现在和为 7 之前”, 那么以首次试验结果为条件也可得到所求概率, 方法如下: 令 F 表示事件“第一次试验中两枚骰子的点数之和为 5”, G 表示事件“第一次试验中点数之和为 7”, H 表示事件“第一次试验中点数之和既不是 5 也不是 7”. 利用全概公式, 有

$$P(E) = P(E|F)P(F) + P(E|G)P(G) + P(E|H)P(H)$$

然而, $P(E|F) = 1, P(E|G) = 0, P(E|H) = P(E)$. 前两个等式是显而易见的, 第三个等式是因为: 第一次结果既不是 5, 也不是 7, 那么这种状况又相当于重新开始了. 也就是说, 试验者将要连续扔两枚骰子直到两枚骰子的点数之和为 5 或者 7, 另一方面, 每次试验都是独立的, 因此, 第一次试验的结果不会对接下来的试验有影响. 因为 $P(F) = 4/36, P(G) = 6/36, P(H) = 26/36$, 这样 $P(E) = \frac{1}{9} + P(E)\frac{13}{18}$, 或者 $P(E) = 2/5$.

读者可能会发现答案很直观, 因为既然“和为 5”出现的概率为 4/36 而“和为 7”出现的概率为 6/36, 那么直观地看出 5 出现在 7 前面的比例为 4 : 6, 因此“和为 5 出现在和为 7 之前”的概率就为 4/10, 事实上的确如此.

这说明了如果 E 和 F 是一次试验中的两个互不相容事件, 那么在独立重复试

验时, 事件 E 在事件 F 之前发生的概率为

$$\frac{P(E)}{P(E)+P(F)}$$

■

例 4i 有 n 种类型的优惠券, 某人在收集优惠券的时候, 每次收集到第 i 种优惠券的概率为 $p_i, \sum_{i=1}^{n} p_i = 1$. 假定各次收集是相互独立的且同分布的. 假设这个人收集了 k 张优惠券, 令 A_i 表示事件“其中至少有一张第 i 种优惠券”, 对于 $i \neq j$, 计算

(a) $P(A_i)$ (b) $P(A_i \cup A_j)$ (c) $P(A_i|A_j)$

解: $P(A_i) = 1 - P(A_i^c) = 1 - P\{没有第i种优惠券\} = 1 - (1-p_i)^k$, 上面利用了每种优惠券的收集都是独立的, 并且收集到的优惠券不是第 i 种优惠券的概率为 $1 - p_i$. 类似地

$$\begin{aligned} P(A_i \cup A_j) &= 1 - P\Big((A_i \cup A_j)^c\Big) \\ &= 1 - P\{既没有第i种优惠券, 也没有第j种优惠券\} \\ &= 1 - (1 - p_i - p_j)^k \end{aligned}$$

此处利用了每种优惠券的收集都是独立的, 且收集到的优惠券既不是第 i 种优惠券也不是第 j 种优惠券的概率为 $1 - p_i - p_j$.

为了计算 $P(A_i|A_j)$, 我们利用等式 $P(A_i \cup A_j) = P(A_i) + P(A_j) - P(A_iA_j)$. 其中, 利用 (a) 和 (b), 可以得到

$$\begin{aligned} P(A_iA_j) &= 1 - (1-p_i)^k + 1 - (1-p_j)^k - [1 - (1-p_i-p_j)^k] \\ &= 1 - (1-p_i)^k - (1-p_j)^k + (1-p_i-p_j)^k \end{aligned}$$

这样,

$$P(A_i|A_j) = \frac{P(A_iA_j)}{P(A_j)} = \frac{1 - (1-p_i)^k - (1-p_j)^k + (1-p_i-p_j)^k}{1-(1-p_j)^k}$$

■

接下来的例子陈述了一个在概率论历史中占有显赫地位的问题, 即著名的**点数问题**[①](problem of the points). 该问题是这样的: 两个赌徒下了注后, 就按照某种方式赌起来, 规定得胜者赢得所有赌注. 但在谁也没得胜之前, 赌博因故中止了. 此时每人都获得了一些“得分”, 那么这些赌本该如何分呢?

这个问题是 1654 年 de Méré 爵士向法国数学家帕斯卡提出的, 爵士当时是一个职业赌徒. 在攻克这一难题的过程中, 帕斯卡提出了这样一个重要的思想: 赌徒赢得的赌本的比例, 取决于如果比赛继续进行下去, 他们各自能取胜的概率. 帕斯卡解决了一些特殊情形, 更主要的是, 他开始与法国著名的数学家费马建立了通信联系, 讨论该问题. 他们之间通信的结果, 不仅完全解决了点数问题, 而且还为解决

① 此处也可解释成赌本分割问题, 更易为大家所理解. ——译者注

很多其他机会游戏问题搭好了框架. 有些人把他们建立通信联系的这一天看作概率论的生日. 他们之间的著名的通信往来, 对于激发欧洲数学家对概率论的兴趣也起了重要的作用. 因为当时一流的数学家都认识帕斯卡和费马. 比如, 在他们建立联系后不久, 荷兰的年轻数学家惠更斯也来到了巴黎, 和他们一起探讨这些问题及其解法. 而且, 人们对这个新领域的兴趣和积极性也迅速高涨起来.

例 4j (点数问题) 假设在独立重复试验中, 每次成功的概率为 p, 失败的概率为 $1-p$. 问在 m 次失败之前已有 n 次成功的概率是多大? 设想 A 和 B 进行这样的赌博: 当试验成功时, A 得 1 分, 试验失败时, B 得 1 分, 如果 A 先得到 n 分那么 A 获胜, 如果 B 先得到 m 分, 那么 B 获胜. 问 A 获胜的概率为多大?

解: 以下将给出两种解答. 第一种是帕斯卡给出的, 第二种是费马给出的.

令 $P_{n,m}$ 表示事件"m 次失败之前已经出现了 n 次成功"的概率, 以第一次的结果为条件, 可得

$$P_{n,m} = pP_{n-1,m} + (1-p)P_{n,m-1} \qquad n \geqslant 1, m \geqslant 1$$

(为什么? 给出理由.) 利用很明显的边界条件 $P_{n,0}=0, P_{0,m}=1$, 该等式能解出 $P_{n,m}$. 与其深入枯燥的细节, 不如来看看费马的解答.

费马论证了, 要使得 n 次成功出现在 m 次失败之前, 那么在前 $m+n-1$ 次试验中, 至少有 n 次成功. (即使在试验 $m+n-1$ 次之前, 赌博就结束了, 我们仍可以假设剩下的试验继续进行.) 事实上, 如果在前 $m+n-1$ 次试验中至少有 n 次成功, 那么至多有 $m-1$ 次失败, 因此 n 次成功必然出现在 m 次失败之前. 另一方面, 如果 $m+n-1$ 次试验中不超过 n 次成功, 那么至少有 m 次失败, 因此, n 次成功不会出现在 m 次失败之前.

因此, 如例 4f 中指出的, $m+n-1$ 次试验中恰好成功 k 次的概率为 $\binom{m+n-1}{k}p^k(1-p)^{m+n-1-k}$, 我们可得所求的 n 次成功出现在 m 次失败之前的概率为

$$P_{n,m} = \sum_{k=n}^{m+n-1} \binom{m+n-1}{k} p^k (1-p)^{m+n-1-k} \qquad \blacksquare$$

下面两个例子是关于赌博问题, 其中例 4k 的分析特别细致. ①

例 4k 假定一共有 r 个玩家, 每个玩家 i 手中有 n_i 个单位的赌资, $n_i > 0, i = 1, \cdots, r$. 在赌博的每一阶段, 随机地选出两个玩家, 让他们进行赌博, 其中赢者从输者那儿拿到一个单位的赌资, 当一个玩家输光了赌资以后, 就退出赌博. 这个过程一直维持到只剩下最后一个人, 此时他拥有全部赌资 $n=\sum_{i=1}^{r} n_i$, 他就是一个胜利者. 现在假定各个阶段之间是相互独立的, 并且参加赌博的任何两个人的能力是相同的, 即在每一阶段的两个参加赌博的玩家具有相同的概率打败对方. 现求玩

① 本节的余下部分为选读.

家 i 成为胜利者的概率.

解: 开始时, 假定 n 个玩家每人只有一个单位的赌资. 考虑玩家 i, 他每次参加赌博, 要么赢一个单位的赌资, 要么失去一个单位的赌资. 他会继续赌下去, 直到他的赌资成为 0 或 n. 由于 n 个玩家的能力是完全相同的, 因此, 在这种情况下, 玩家 i 成为最后胜利者的概率为 $1/n$.

现在假定 n 个玩家分成 r 个组, 第 i 个组内含 n_i 个玩家, $i=1,\cdots,r$. 这样, 第 i 组的玩家成为胜利者的概率为 n_i/n. 如果把 i 组的玩家成为胜利者看成 i 组的胜利, 在各阶段各组的赌资就是组内各玩家赌资的总和, 将赌博看成团体间的赌博. 这样, 在初始状态各参赌团体的赌资为 $n_i, i=1,\cdots,r$. 每一阶段的赌博, 参赌的团体的赌资增加一个单位或减少一个单位, 显然, 第 i 个参赌团体胜利的概率就是 $P_i=n_i/n$. 由分析可以看出, 这个结果与各阶段的参赌团体的选择是无关的. ■

接下来的例子是著名的"赌徒破产问题".

例 41(赌徒破产问题) 两个赌徒, 就连续抛掷一枚硬币的结果进行打赌. 对于每一次抛掷, 如果是正面朝上, B 将支付给 A 一元, 如果是反面朝上, A 将付给 B 一元. 一直这样下去, 直到某一方钱输光. 假定连续抛掷硬币是独立的, 且每次的结果正面朝上的概率为 p, 假定开始时 A 有 i 元, B 有 $N-i$ 元, 问 A 最后能赢得所有钱的概率是多大.

解: 令 E 表示事件"开始时 A 有 i 元, B 有 $N-i$ 元, 而 A 最后赢得所有钱". 很显然它和 A 最初的钱数是有关的, 记 $P_i=P(E)$. 以第一次掷硬币的结果为条件, 令 H 表示事件"第一次为正面朝上", 这样得到 $P(E)$ 的表达式如下:

$$P_i=P(E)=P(E|H)P(H)+P(E|H^c)P(H^c)=pP(E|H)+(1-p)P(E|H^c)$$

现在, 假定第一次硬币为正面朝上, 第一次打赌结束后的状态是: A 有了 $i+1$ 元, 而 B 有 $N-(i+1)$ 元. 因为随后的抛掷都同前面独立并且正面朝上的概率都为 p, 因此, 从该时刻开始, A 赢得所有钱的概率等同于这种情形: 一开始 A 有 $i+1$ 元而 B 有 $N-(i+1)$ 元. 因此 $P(E|H)=P_{i+1}$. 类似地, $P(E|H^c)=P_{i-1}$ 这样, 令 $q=1-p$, 可以得到

$$P_i=pP_{i+1}+qP_{i-1} \qquad i=1,2,\cdots,N-1 \tag{4.2}$$

利用明显的边界条件 $P_0=0$ 和 $P_N=1$, 就可以求解 (4.2) 式. 由于 $p+q=1$, 上式等价于

$$pP_i+qP_i=pP_{i+1}+qP_{i-1}$$

或者

$$P_{i+1}-P_i=\frac{q}{p}(P_i-P_{i-1}) \qquad i=1,2,\cdots,N-1 \tag{4.3}$$

这样, 由 $P_0=0$, 利用 (4.3) 式可以得到

$$P_2-P_1=\frac{q}{p}(P_1-P_0)=\frac{q}{p}P_1$$

$$P_3 - P_2 = \frac{q}{p}(P_2 - P_1) = \left(\frac{q}{p}\right)^2 P_1$$

$$\vdots$$

$$P_i - P_{i-1} = \frac{q}{p}(P_{i-1} - P_{i-2}) = \left(\frac{q}{p}\right)^{i-1} P_1 \tag{4.4}$$

$$\vdots$$

$$P_N - P_{N-1} = \frac{q}{p}(P_{N-1} - P_{N-2}) = \left(\frac{q}{p}\right)^{N-1} P_1$$

将 (4.4) 式中前 $i-1$ 个等式累加, 可以得到

$$P_i - P_1 = P_1\left[\left(\frac{q}{p}\right) + \left(\frac{q}{p}\right)^2 + \cdots + \left(\frac{q}{p}\right)^{i-1}\right]$$

或者

$$P_i = \begin{cases} \dfrac{1-(q/p)^i}{1-(q/p)} P_1 & 如果 \dfrac{q}{p} \neq 1 \\ iP_1 & 如果 \dfrac{q}{p} = 1 \end{cases}$$

再利用事实 $P_N = 1$, 可以得到

$$P_1 = \begin{cases} \dfrac{1-(q/p)}{1-(q/p)^N} & 如果 p \neq \dfrac{1}{2} \\ \dfrac{1}{N} & 如果 p = \dfrac{1}{2} \end{cases}$$

因此

$$P_i = \begin{cases} \dfrac{1-(q/p)^i}{1-(q/p)^N} & 如果 p \neq \dfrac{1}{2} \\ \dfrac{i}{N} & 如果 p = \dfrac{1}{2} \end{cases} \tag{4.5}$$

令 Q_i 表示开始 A 有 i 元, B 有 $N-i$ 元, 最后 B 赢得所有钱的概率, 那么, 由对称性, 只需将 p 替换为 q, i 替换为 $N-i$, 我们可以看出

$$Q_i = \begin{cases} \dfrac{1-(p/q)^{N-i}}{1-(p/q)^N} & 如果 q \neq \dfrac{1}{2} \\ \dfrac{N-i}{N} & 如果 q = \dfrac{1}{2} \end{cases}$$

而且, 因为 $q = 1/2$ 等价于 $p = 1/2$, 因此当 $q \neq 1/2$ 时, 有

$$\begin{aligned} P_i + Q_i &= \frac{1-(q/p)^i}{1-(q/p)^N} + \frac{1-(p/q)^{N-i}}{1-(p/q)^N} \\ &= \frac{p^N - p^N(q/p)^i}{p^N - q^N} + \frac{q^N - q^N(p/q)^{N-i}}{q^N - p^N} \\ &= \frac{p^N - p^{N-i}q^i - q^N + q^i p^{N-i}}{p^N - q^N} = 1 \end{aligned}$$

当 $p = q = 1/2$ 时结论仍成立, 因此我们有

$$P_i + Q_i = 1$$

上式表明, A 和 B 中某一人将赢得所有钱的概率为 1; 或者说, A 的钱总在 1 与 $N-1$ 之间而赌博无休止地进行下去的概率为 0(读者必须注意, 这场赌博有三个可能结果, 而不是两个, 即或者 A 胜, 或者 B 胜, 或者谁也不胜. 但我们刚才证明了最后一种结果的概率为 0.)

现在给上述结论以数量上的说明. 若开始 A 有 5 元, 而 B 有 10 元, 则当 $p=1/2$ 时, A 得胜的概率为 1/3, 而当 $p=0.6$ 时, A 得胜的概率猛增为

$$\frac{1-\left(\frac{2}{3}\right)^5}{1-\left(\frac{2}{3}\right)^{15}} \approx 0.87$$

赌徒破产问题还有一种特殊情形, 称为**赌博持续时间问题** (duration of play). 这个问题是 1657 年法国数学家费马向荷兰数学家克里斯第安 · 惠更斯提出的, 后来被惠更斯解决. 惠更斯解决的版本是这样的, 设 A 和 B 每人有 12 枚硬币, 他们以抛掷 3 个骰子的方法赌这些钱: 若点数为 11(不管谁掷骰子都可以), 则 A 给 B 一枚硬币, 如果点数为 14, 则 B 给 A 一枚硬币. 谁先赢得所有硬币谁就获胜. 因为 $P\{\text{点数为 }11\} = 27/216$ 及 $P\{\text{点数为 }14\} = 15/216$, 由例 4h 可以看出, 对 A 而言, 这正是 $p = 12/45, i = 12, N = 24$ 情形下的赌徒破产问题 (例 4k). 一般的赌徒破产问题由数学家詹姆士 · 伯努利解决, 其结果发表于 1713 年 (他去世后的第 8 年).

作为赌徒破产问题的应用, 讨论如下的药品试验问题. 设为治疗某种疾病, 正在研制两种新药. 新药 i 的治愈率为 $P_i, i=1,2$. 然而, P_i 为未知的. 我们希望知道 $P_i > P_j$ 或 $P_j > P_i$. 试验是成对地、有序地进行的. 对于各对病人, 其中一人施以药 1, 另一人用药 2. 当其中一种药的治愈人数超过另一种药的治愈人数的一定数量时, 试验就停止. 令

$$X_j = \begin{cases} 1 & \text{第}j\text{对病人中用第 1 种药品者治愈了他的疾病} \\ 0 & \text{其他} \end{cases}$$

$$Y_j = \begin{cases} 1 & \text{第}j\text{对病人中用第 2 种药品者治愈了他的疾病} \\ 0 & \text{其他} \end{cases}$$

设 M 是事先确定的正整数, 试验在第 N 次时停止, 其中 N 是使下列两个等式中某一个第一次成立时的那个 n 的值:

$$X_1 + \cdots + X_n - (Y_1 + \cdots + Y_n) = M$$

或

$$X_1 + \cdots + X_n - (Y_1 + \cdots + Y_n) = -M$$

若 n 使第一个等式成立, 就下结论 $P_1 > P_2$, 若 n 使第二个等式成立, 则下结论

$P_1 < P_2$.

为了验证这个方法的好坏，我们需要知道利用这个方法导致作出错误结论的概率. 即在假设 $P_1 > P_2$ 是一组药的治愈率的真值的条件下，作出"$P_2 > P_1$"的错误决定的概率. 注意到每一次试验的结果：当药品 1 有效，而药品 2 无效时，这个累计差会增加 1，相应的概率为 $P_1(1-P_2)$. 但如果药品 2 有效而药品 1 无效时，这个累计差会减少 1，其相应的概率为 $P_2(1-P_1)$. 当两种药的疗效相同时，这个累计差保持不变，其相应概率为 $P_1P_2+(1-P_1)(1-P_2)$. 如果我们只考虑累计差变化的那些试验，则累计差以概率

$$P = P\{\text{累计差增1}|\text{累计差增 1 或减 1}\} = \frac{P_1(1-P_2)}{P_1(1-P_2)+P_2(1-P_1)}$$

增 1，以概率

$$1-P = \frac{P_2(1-P_1)}{P_1(1-P_2)+P_2(1-P_1)}$$

减 1. 因此，作出判断"$P_2 > P_1$"的概率等于赌徒累计赢 M 元钱之前累计输 M 元钱的概率 (赌徒每次赢的概率为 p，每次输赢为 1 元钱). 在公式 (4.5) 中，令 $i = M, N = 2M$，这个概率为

$$P\{\text{下结论}P_2 > P_1\} = 1 - \frac{1-\left(\frac{1-P}{P}\right)^M}{1-\left(\frac{1-P}{P}\right)^{2M}} = 1 - \frac{1}{1+\left(\frac{1-P}{P}\right)^M} = \frac{1}{1+\gamma^M}$$

其中 $\gamma = \frac{P}{1-P} = \frac{P_1(1-P_2)}{P_2(1-P_1)}$，例如，$P_1 = 0.6, P_2 = 0.4$，则当 $M = 5$ 时，作出不正确判断的概率为 0.017，当 $M = 10$ 时，这个概率降到 0.0003. ■

现在我们介绍一种解决非概率问题的概率方法. 设有一个集合，我们希望知道在这个集合中是否至少有一个元素具有某一特征. 例如，在一群鸡中，我们希望知道是否有鸡患禽流感. 我们的方法是随机地从这一集合中取一元素，抽取的方法是使得这个集合中的每一个元素被抽取到的概率都大于 0. 现在考虑抽到的元素不具有该特征的概率. 若这个概率为 1，则在这个集合中没有元素有这个特征. 若这个概率小于 1，则在这个集合中至少有一元素具有这个特征.

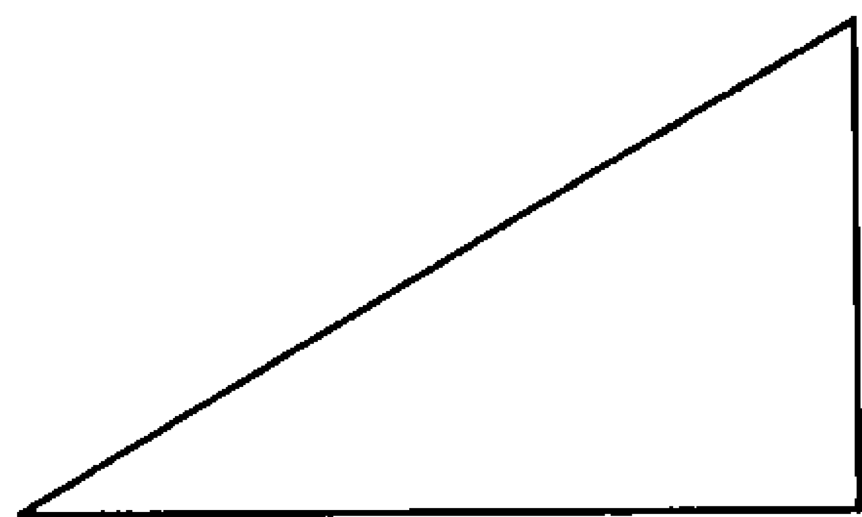


图3.3 3个顶点的完全图

本节最后一个例子就是应用这个技术.

例 4m 首先给出 n 个顶点的完全图的定义：在平面上有 n 个点 (称为顶点)，用 $\binom{n}{2}$ 条线段 (称为边) 将这些点联结起来. 如图 3.3 表示了 3 个顶点的完全图. 假设 n 个顶点的完全图的每条边染成了红色或者蓝色. 一个有趣的问题是：对于给定的整数 k，是否存在一个染色方法，使得该图上任意给定的 k 个顶点，其相应的 $\binom{k}{2}$ 条边不是同一颜色. 通过概率讨论可以证明，如果 n 不是太大，答案是肯定的.

讨论如下：假设每条边独立等可能地染成红色或蓝色. 也即, 每条边为红色的概率为 1/2. 将这 $\binom{n}{k}$ 个 k 个顶点所组成的子集编号, 定义事件 $E_i, i=1,\cdots,\binom{n}{k}$ 如下：

$$E_i = \{\text{第}i\text{个}k\text{顶点集合的所有边颜色相同}\}$$

这样, 由于 k 顶点集合的 $\binom{k}{2}$ 条边的每条都等可能地染成红色或蓝色, 因此, 它们颜色相同的概率为

$$P(E_i) = 2\left(\frac{1}{2}\right)^{k(k-1)/2}$$

由

$$P\left(\bigcup_i E_i\right) \leqslant \sum_i P(E_i) \qquad (\text{布尔不等式})$$

我们可得“至少存在一个 k 点集合, 其所有边颜色相同”的概率为 $P\left(\bigcup_i E_i\right)$, 满足

$$P\left(\bigcup_i E_i\right) \leqslant \binom{n}{k}\left(\frac{1}{2}\right)^{k(k-1)/2-1}$$

因此, 如果 $\binom{n}{k}\left(\frac{1}{2}\right)^{k(k-1)/2-1} < 1$, 或者, 等价于 $\binom{n}{k} < 2^{k(k-1)/2-1}$, 则 $\binom{n}{k}$ 个 k 点集合里, “至少存在一个 k 点集合, 其所有的边颜色都相同”的概率小于 1. 因此, 在前述 n 和 k 的条件下, “没有一个 k 点集合, 其所有边颜色相同”的概率为正数, 这意味着至少存在一种染色方法, 使得对任意 k 顶点集合, 其所有的边染色不全相同. ■

注释 (a) 上面的论证列出了关于 n, k 的条件, 在这样条件之下, 存在一种涂颜色的方法即可满足所要求的性质. 但它并没有告诉我们如何涂颜色, 使得所涂的颜色满足所要求的性质. (当然, 可以随机地涂色, 然后检查所涂的颜色是否满足所要求的性质, 若不成, 再重复一次, 直到成功为止.)

(b) 将概率引进那些纯粹是确定问题的方法称为概率化方法 (probabilistic method①). 此方法的其他例子在理论习题 24 以及第 7 章的例 2t 和例 2u 中给出.

3.5 $P(\cdot|F)$ 为概率

条件概率满足普通概率的所有性质. 命题 5.1 证明了条件概率 $P(E|F)$ 满足概率的三条公理.

---

① 参见 N. Alon, J. Spencer 及 P. Erdos, *The Probabilistic Method* (纽约: John Wiley & Sons, Inc., 1992).

命题 5.1 (a) $0 \leqslant P(E|F) \leqslant 1$; (b) $P(S|F) = 1$; (c) 若 $E_i, i = 1, 2, \cdots$ 为互不相容事件列, 则

$$P\left(\bigcup_{i=1}^{\infty} E_i|F\right) = \sum_{i=1}^{\infty} P(E_i|F)$$

证明: 为了证明 (a), 我们只要证明 $0 \leqslant P(EF)/P(F) \leqslant 1$ 即可. 不等式左边是显然的, 而不等式的右边成立是因为 $EF \subset F$ 成立意味着 $P(EF) \leqslant P(F)$. (b) 成立是因为

$$P(S|F) = \frac{P(SF)}{P(F)} = \frac{P(F)}{P(F)} = 1$$

(c) 成立是因为

$$P\left(\bigcup_{i=1}^{\infty} E_i|F\right) = \frac{P\left(\left(\bigcup_{i=1}^{\infty} E_i\right)F\right)}{P(F)} = \frac{P\left(\bigcup_{i=1}^{\infty} E_iF\right)}{P(F)} \qquad \text{因为} \left(\bigcup_{i=1}^{\infty} E_i\right)F = \bigcup_{i=1}^{\infty} E_iF$$

$$= \frac{\sum_{i=1}^{\infty} P(E_iF)}{P(F)} = \sum_{i=1}^{\infty} P(E_i|F)$$

其中倒数第二个等式成立是因为 $E_iE_j = \varnothing$, 这意味着 $E_iFE_jF = \varnothing$. □

如果我们定义 $Q(E) = P(E|F)$, 根据命题 5.1, $Q(E)$ 可认为是关于 S 中事件的概率函数. 因此, 前面证明的关于概率的命题它都满足. 举例说, 我们有

$$Q(E_1 \cup E_2) = Q(E_1) + Q(E_2) - Q(E_1E_2)$$

或者等价地,

$$P(E_1 \cup E_2|F) = P(E_1|F) + P(E_2|F) - P(E_1E_2|F)$$

而且, 如果我们再定义条件概率 $Q(E_1|E_2)$ 如下: $Q(E_1|E_2) = Q(E_1E_2)/Q(E_2)$, 根据 (3.1) 式可得

$$Q(E_1) = Q(E_1|E_2)Q(E_2) + Q(E_1|E_2^c)Q(E_2^c) \tag{5.1}$$

由于

$$Q(E_1|E_2) = \frac{Q(E_1E_2)}{Q(E_2)} = \frac{P(E_1E_2|F)}{P(E_2|F)} = \frac{P(E_1E_2F)/P(F)}{P(E_2F)/P(F)} = P(E_1|E_2F)$$

我们可看出 (5.1) 式等价于

$$P(E_1|F) = P(E_1|E_2F)P(E_2|F) + P(E_1|E_2^cF)P(E_2^c|F)$$

例 5a 考虑例 3a, 保险公司认为人可以分为不同的两类, 一类是易出事故的, 另一类是比较谨慎的. 在任意给定的一年内, 易出事故的人将发生事故的概率为 0.4, 而对比较谨慎的人来说, 此概率为 0.2. 若已知某新保险客户在第一年已经出过一次事故, 问他在保险有效的第二年又出一次事故的条件概率是多大?

解: 如果令 A 表示"该保险客户是易出事故的人"这一事件, 而 $A_i, i=1,2$ 表示"他在第 i 年出一次事故". 那么, 以他是不是易出事故的人为条件, 可以算出所求概率 $P(A_2|A_1)$ 如下:

$$P(A_2|A_1)=P(A_2|AA_1)P(A|A_1)+P(A_2|A^{c}A_1)P(A^{c}|A_1)$$

而

$$P(A|A_1)=\frac{P(A_1A)}{P(A_1)}=\frac{P(A_1|A)P(A)}{P(A_1)}$$

但是, 例 3a 中已经假设 $P(A)=3/10$, 且算出了 $P(A_1)=0.26$, 因此

$$P(A|A_1)=\frac{0.4\times 0.3}{0.26}=\frac{6}{13}$$

从而

$$P(A^{c}|A_1)=1-P(A|A_1)=\frac{7}{13}$$

由于 $P(A_2|AA_1)=0.4$ 以及 $P(A_2|A^{c}A_1)=0.2$, 我们得到

$$P(A_2|A_1)=0.4\times\frac{6}{13}+0.2\times\frac{7}{13}\approx 0.29$$

■

例 5b 一只母猩猩生了一只幼猩猩, 但是, 却不能断定两只公猩猩究竟哪一只是父亲. 在进行基因分析之前, 有迹象表明第一只公猩猩为父亲的概率为 p, 第二只为父亲的概率为 $1-p$. 从这三只猩猩身上获得的 DNA 表明, 对于一个特殊的基因组, 母猩猩具有基因对 (A,A), 第一只公猩猩具有基因对 (a,a), 而第二只公猩猩具有基因对 (A,a). 幼猩猩出生后, 具有基因对 (A,a), 第一只公猩猩是父亲的概率是多大?

解: 令所有概率都是以"母猩猩有基因对 (A,A), 第一只公猩猩具有基因对 (a,a), 第二只公猩猩具有基因对 (A,a)"为条件的条件概率. 又令 M_i 表示第 i 只公猩猩为父亲这一事件, $i=1,2$. 令 $B_{\mathrm{A,a}}$ 表示幼猩猩具有基因对 (A,a) 这一事件, 那么如下可得到 $P(M_1|B_{\mathrm{A,a}})$:

$$\begin{aligned}P(M_1|B_{\mathrm{A,a}})&=\frac{P(M_1B_{\mathrm{A,a}})}{P(B_{\mathrm{A,a}})}=\frac{P(B_{\mathrm{A,a}}|M_1)P(M_1)}{P(B_{\mathrm{A,a}}|M_1)P(M_1)+P(B_{\mathrm{A,a}}|M_2)P(M_2)}\\&=\frac{1\times p}{1\times p+1/2\times(1-p)}=\frac{2p}{1+p}\end{aligned}$$

由于当 $p<1$ 时 $2p/(1+p)>p$, 我们可以看出幼猩猩的基因对为 (A,a) 这一信息增加了第一只公猩猩为父亲的概率. 这点很直观, 因为相对于 M_2 来说, M_1 成

立的条件下, 幼猩猩的基因对为 (A,a) 的可能性更大 (各自的条件概率分别为 1 和 1/2). ■

下面的例子研究游程的理论中的一个问题.

例 5c 设有一个独立重复试验序列, 每次成功的概率为 p, 失败的概率为 $q = 1-p$. 我们希望计算长度为 n 的成功游程先于长度为 m 的失败游程的概率.

解: 记 E 为事件"长度为 n 的成功游程先于长度为 m 的失败游程". 以第一次试验结果为条件, 得到

$$P(E) = pP(E|H) + qP(E|H^c) \tag{5.2}$$

其中 H 表示第一次试验成功. 现在假定第一次试验成功, 为了使得长度为 n 的成功游程先出现, 我们希望接着 $n-1$ 次试验均为成功. 现在令 F 表示事件"第 2 次到第 n 次试验均成功". 以 F 为条件, 我们得到

$$P(E|H) = P(E|FH)P(F|H) + P(E|F^cH)P(F^c|H) \tag{5.3}$$

显然, $P(E|FH) = 1$. 另一方面, 若 F^cH 发生, 说明第一次试验成功, 但在后面的 $n-1$ 次试验中, 至少有一次试验失败. 此时, 前面的成功已经失去作用, 相当于从失败开始, 因此 $P(E|F^cH) = P(E|H^c)$, 由试验的相互独立性, 可知 $P(F|H) = P(F) = p^{n-1}$. 因此, 由 (5.3) 可知

$$P(E|H) = p^{n-1} + (1-p^{n-1})P(E|H^c) \tag{5.4}$$

现在考虑 $P(E|H^c)$. 令 G 表示事件"第 2 次试验直到第 m 次试验均失败". 我们得到

$$P(E|H^c) = P(E|GH^c)P(G|H^c) + P(E|G^cH^c)P(G^c|H^c) \tag{5.5}$$

GH^c 表示前 m 次试验均失败, 故 $P(E|GH^c) = 0$. 当 G^cH^c 发生时, 所有以前的失败已经失去作用, 相当于从成功开始, 即 $P(E|G^cH^c) = P(E|H)$. 又由于 $P(G^c|H^c) = P(G^c) = 1-q^{m-1}$, (5.5) 化为

$$P(E|H^c) = (1-q^{m-1})P(E|H) \tag{5.6}$$

由 (5.4) 和 (5.6) 可得

$$P(E|H) = \frac{p^{n-1}}{p^{n-1}+q^{m-1}-p^{n-1}q^{m-1}}, \quad P(E|H^c) = \frac{(1-q^{m-1})p^{n-1}}{p^{n-1}+q^{m-1}-p^{n-1}q^{m-1}}$$

最后得到

$$\begin{aligned} P(E) &= pP(E|H) + qP(E|H^c) \\ &= \frac{p^n + qp^{n-1}(1-q^{m-1})}{p^{n-1}+q^{m-1}-p^{n-1}q^{m-1}} = \frac{p^{n-1}(1-q^m)}{p^{n-1}+q^{m-1}-p^{n-1}q^{m-1}} \end{aligned} \tag{5.7}$$

由对称性, 可知长度为 m 的失败游程先于长度为 n 的成功游程的概率也可用 (5.7)

式进行计算, 不过公式中 m 和 n 对换, p 和 q 对换.

$$P\{\text{长度为 } m \text{ 的失败游程出现在长度为 } n \text{ 的成功游程之前}\}$$
$$= \frac{q^{m-1}(1-p^n)}{q^{m-1}+p^{n-1}-q^{m-1}p^{n-1}} \tag{5.8}$$

由于 (5.7) 和 (5.8) 之和为 1, 可知长度为 n 的成功游程和长度为 m 的失败游程终有一个会出现.

举例, 掷一枚均匀的硬币, 长度为 2 的正面游程先于长度为 3 的反面游程的概率为 7/10, 长度为 2 的正面游程先于长度为 4 的反面游程的概率为 5/6. ■

例 5d 再次研究配对问题 (第 2 章例 5m), 这次运用条件概率解答问题.

例 5d 在一次聚会上, n 个人摘下他们的帽子, 然后把这些帽子混合在一起, 每人再随机选择一顶. 如某个人选中了他自己的帽子, 我们就说出现了一个配对. 以下事件概率是多大?

(a) 没有配对. (b)恰有 k 个配对.

解: (a) 令 E 表示"没有配对"这一事件, 它显然与 n 有关, 因此可记 $P_n = P(E)$. 以第一个人是否选中自己的帽子 (分别记为 M 和 M^c) 为条件, 有

$$P_n = P(E) = P(E|M)P(M) + P(E|M^c)P(M^c)$$

易知 $P(E|M) = 0$, 因此

$$P_n = P(E|M^c)\frac{n-1}{n} \tag{5.9}$$

$P(E|M^c)$ 是在已知 $n-1$ 个人中有一个特殊的人 (此人的帽子已被第一人选走) 必定选不到自己帽子的条件下, 这 $n-1$ 个人选 $n-1$ 顶帽子没有配对的概率. 此处有两种互不相容的选取方式: 该特殊的人没有选中第一人的帽子且其余的人中也没有配对; 该特殊的人选中了第一人的帽子, 且其余的人中也没有配对. 前者的概率正是 P_{n-1}(此时可把该特殊的人理解成第一人), 而后者的概率为 $[1/(n-1)]P_{n-2}$. 这样, 我们得到

$$P(E|M^c) = P_{n-1} + \frac{1}{n-1}P_{n-2}$$

于是由 (5.9) 式可得

$$P_n = \frac{n-1}{n}P_{n-1} + \frac{1}{n}P_{n-2}$$

或等价地有

$$P_n - P_{n-1} = -\frac{1}{n}(P_{n-1} - P_{n-2}) \tag{5.10}$$

但是, 由于 P_n 表示 n 个人在他们的帽子中随机地选一顶但没有一人选中自己的帽

子的概率, 我们有 $P_1=0, P_2=1/2$, 从而由 (5.10) 式可得:

$$P_3-P_2=-\frac{P_2-P_1}{3}=-\frac{1}{3!} \quad 或 \quad P_3=\frac{1}{2!}-\frac{1}{3!}$$

$$P_4-P_3=-\frac{P_3-P_2}{4}=\frac{1}{4!} \quad 或 \quad P_4=\frac{1}{2!}-\frac{1}{3!}+\frac{1}{4!}$$

一般地, 我们有

$$P_n=\frac{1}{2!}-\frac{1}{3!}+\frac{1}{4!}-\cdots+\frac{(-1)^n}{n!}$$

(b) 为了计算正好有 k 个配对的概率, 先考虑固定的某 k 个人, 只有他们选中自己的帽子的概率为

$$\frac{1}{n}\cdot\frac{1}{n-1}\cdots\frac{1}{n-(k-1)}P_{n-k}=\frac{(n-k)!}{n!}P_{n-k}$$

其中 P_{n-k} 是已知 k 个人选中自己的帽子, 其余 $n-k$ 个人在他们自己的帽子中选取而没有配对的条件概率, 再因这 k 个人有 $\binom{n}{k}$ 种选法, 故正好有 k 个配对的概率为

$$\frac{P_{n-k}}{k!}=\frac{\dfrac{1}{2!}-\dfrac{1}{3!}+\cdots+\dfrac{(-1)^{n-k}}{(n-k)!}}{k!}$$

■

概率论的另一个重要概念是事件的条件独立性. 我们称事件 E_1 和 E_2 对于给定的事件 F 是**条件独立的** (conditionally independent), 如果已知 F 发生的条件下, E_1 发生的概率不因 E_2 是否发生而改变. 确切地说, 称 E_1 与 E_2 在给定 F 发生之下是条件独立的, 如果

$$P(E_1|E_2F)=P(E_1|F) \tag{5.11}$$

或等价地,

$$P(E_1E_2|F)=P(E_1|F)P(E_2|F) \tag{5.12}$$

条件独立的概念容易推广到两个以上事件的情形, 我们把它留作习题.

读者会发现, 条件独立性的概念在 例 5a 中已经用过了. 在那里, 我们假定: 在已知保险客户是否为易出事故的人的情况下, "他在第 i 年, $i=1,2,\cdots$, 出一次事故" 这些事件是条件独立的. 下一个例题, 有时也称为拉普拉斯继承准则, 进一步解释条件独立的概念.

例 5e (拉普拉斯继承准则) 盒中有 $k+1$ 枚不均匀的硬币, 抛掷第 i 枚硬币时, 其正面朝上的概率为 $i/k, i=0,1,\cdots,k$. 从盒子中随机取出一枚硬币, 并重复地抛掷, 若前 n 次抛掷结果都为正面朝上, 问第 $n+1$ 次结果仍为正面朝上的概率是多大?

解: 令 C_i 表示"开始取出的是第 i 枚硬币"这一事件, $i=0,1,\cdots,k$, F_n 表示"前 n 次结果都为正面朝上", H 表示"第 $n+1$ 次抛掷出现正面朝上". 所求概率

$P(H|F_n)$ 为

$$P(H|F_n)=\sum_{i=0}^{k}P(H|F_nC_i)P(C_i|F_n)$$

现已知取出的是第 i 枚硬币, 假设各次抛掷的结果是条件独立的, 每次出现正面朝上的概率为 i/k, 于是有

$$P(H|F_nC_i)=P(H|C_i)=\frac{i}{k}$$

而且,

$$P(C_i|F_n)=\frac{P(C_iF_n)}{P(F_n)}=\frac{P(F_n|C_i)P(C_i)}{\sum\limits_{j=0}^{k}P(F_n|C_j)P(C_j)}=\frac{(i/k)^n[1/(k+1)]}{\sum\limits_{j=0}^{k}(j/k)^n[1/(k+1)]}$$

因此有

$$P(H|F_n)=\frac{\sum\limits_{i=0}^{k}(i/k)^{n+1}}{\sum\limits_{j=0}^{k}(j/k)^n}$$

但当 k 充分大时, 可利用积分近似

$$\frac{1}{k}\sum_{i=0}^{k}\left(\frac{i}{k}\right)^{n+1}\approx\int_0^1x^{n+1}\,\mathrm{d}x=\frac{1}{n+2}\quad\frac{1}{k}\sum_{j=0}^{k}\left(\frac{j}{k}\right)^{n}\approx\int_0^1x^{n}\,\mathrm{d}x=\frac{1}{n+1}$$

故对很大的 k 有

$$P(H|F_n)\approx\frac{n+1}{n+2}$$ ■

例 5f (序贯地补充信息) 假设有 n 个互不相容且完全的假设, 其初始概率 [有时也称为先验 (prior)概率] 为 $P(H_i)$, $\sum_{i=1}^{n}P(H_i)=1$. 现在假设得到信息: 事件 E 发生, 那么 H_i 成立的条件概率为 [有时称为 H_i 的后验(posterior)概率]:

$$P(H_i|E)=\frac{P(E|H_i)P(H_i)}{\sum\limits_{j}P(E|H_j)P(H_j)}\tag{5.13}$$

现在, 假设我们首先知道 E_1 发生, 然后 E_2 发生. 那么, 如果仅仅得知第一条信息时, H_i 为真假设的条件概率为:

$$P(H_i|E_1)=\frac{P(E_1|H_i)P(H_i)}{P(E_1)}=\frac{P(E_1|H_i)P(H_i)}{\sum\limits_{j}P(E_1|H_j)P(H_j)}$$

而如果得知两条信息时, H_i 为真假设的条件概率 $P(H_i|E_1E_2)$ 可以如下计算:

$$P(H_i|E_1E_2) = \frac{P(E_1E_2|H_i)P(H_i)}{\sum_j P(E_1E_2|H_j)P(H_j)}$$

然而, 或许有人疑惑, 可否这样计算 $P(H_i|E_1E_2)$: 利用 (5.13) 式的右边, 将 E 替换为 E_2, 将 $P(H_j)$ 替换为 $P(H_j|E_1), j=1,\cdots,n$. 也即, 将 $P(H_j|E_1), j\geqslant 1$ 作为先验概率, 将 E_2 作为新近得到的信息, 然后利用 (5.13) 式来计算后验概率?

解: 上述算法是合理的, 条件是: 对每一个 $j=1,\cdots,n$, 在给定 H_j 下, 事件 E_1 和 E_2 是条件独立的. 如果这样, 那么

$$P(E_1E_2|H_j) = P(E_2|H_j)P(E_1|H_j) \qquad j=1,\cdots,n$$

这样,

$$\begin{aligned}
P(H_i|E_1E_2) &= \frac{P(E_2|H_i)P(E_1|H_i)P(H_i)}{P(E_1E_2)} = \frac{P(E_2|H_i)P(E_1H_i)}{P(E_1E_2)} \\
&= \frac{P(E_2|H_i)P(H_i|E_1)P(E_1)}{P(E_1E_2)} = \frac{P(E_2|H_i)P(H_i|E_1)}{Q(1,2)}
\end{aligned}$$

其中 $Q(1,2)=P(E_1E_2)/P(E_1)$, 由于上式对所有 i 都成立, 我们将上式对 i 求和得到

$$1 = \sum_{i=1}^{n} P(H_i|E_1E_2) = \sum_{i=1}^{n} \frac{P(E_2|H_i)P(H_i|E_1)}{Q(1,2)}$$

说明

$$Q(1,2) = \sum_{i=1}^{n} P(E_2|H_i)P(H_i|E_1)$$

这样可得结果为

$$P(H_i|E_1E_2) = \frac{P(E_2|H_i)P(H_i|E_1)}{\sum_{i=1}^{n} P(E_2|H_i)P(H_i|E_1)}$$

举例来说, 假设有两枚硬币, 选择一枚抛掷, 令 H_i 表示选中了第 i 枚硬币, $i=1,2$, 并假设选中第 i 枚硬币后抛掷, 正面朝上的概率为 $p_i, i=1,2$. 令 E_j 表示"对于选中的硬币的第 j 次的抛掷结果". 抛掷以后, 即 E_1 发生以后, 只需将 $P(H_i)$ 进行修正, 得到 $P(H_i|E_1)$. 若以后还有新的试验结果 E_2, 此时只需将 $P(H_i|E_1)$ 进行修正, 得到 $P(H_i|E_1E_2)$, 将 $P(H_i|E_1E_2)$ 重新写成 $P(H_i)$, 即忘掉它的历史. 每次得到新的试验结果, 只需将修正了的 $P(H_i)$ 再一次进行修正即可, 而不必考虑修正的历史. ■

小 结

对于任意事件 E 和 F, 已知 F 发生的条件下, E 发生的条件概率记为 $P(E|F)$,

定义为如下：

$$P(E|F)=\frac{P(EF)}{P(F)}$$

等式 $P(E_1E_2\cdots E_n)=P(E_1)P(E_2|E_1)\cdots P(E_n|E_1\cdots E_{n-1})$ 称为概率的**乘法规则**.

一个有用的等式 $P(E)=P(E|F)P(F)+P(E|F^c)P(F^c)$ 可用来通过以 F 是否发生为条件计算 $P(E)$.

$P(H)/P(H^c)$ 称为事件 H 的**优势**. 等式

$$\frac{P(H|E)}{P(H^c|E)}=\frac{P(H)P(E|H)}{P(H^c)P(E|H^c)}$$

说明了当得到一个新的证据 E 后, H 的优势等于原来的优势值乘以当 H 成立时新证据发生的概率与 H 不成立时新证据发生的概率的比值.

令 $F_i, i=1,\cdots,n$ 为互不相容事件列, 且它们的并为整个样本空间, 等式

$$P(F_j|E)=\frac{P(E|F_j)P(F_j)}{\sum_{i=1}^{n}P(E|F_i)P(F_i)}$$

称为贝叶斯公式. 如果事件 $F_i, i=1,\cdots,n$ 为一组假设, 那么贝叶斯公式说明了如何计算当新证据成立时, 这些假设成立的条件概率.

如果 $P(EF)=P(E)P(F)$, 那么我们称事件 E 和 F 是**独立**的. 该等式等价于 $P(E|F)=P(E)$ 或 $P(F|E)=P(F)$. 也即, 如果知道其中之一的发生并不影响另一个的发生的概率, 那么 E 和 F 独立.

事件 $E_1,\cdots,E_n$ 称为独立的, 如果对任何子集 $E_{i_1},\cdots,E_{i_r}$, 有

$$P(E_{i_1}\cdots E_{i_r})=P(E_{i_1})\cdots P(E_{i_r})$$

对于任一给定事件 F, $P(E|F)$ 可以认为是样本空间的事件 E 的概率函数.

习　题

1. 掷两枚均匀骰子, 求已知两枚骰子点数不同的条件下, 至少有一枚点数为 6 的条件概率.

2. 掷两枚均匀骰子, 求给定两枚骰子点数之和为 i 的条件下, 第一枚点数为 6 的条件概率. 计算 i 取值从 2 到 12 的所有情况.

3. 利用公式 (2.1), 计算在一手桥牌里, 南北两家持有总共 8 张黑桃的条件下, 东家持有 3 张黑桃的条件概率.

4. 掷两枚均匀骰子, 在点数和为 $i, i=2,3,\cdots,12$ 的条件下, 至少有一枚点数为 6 的条件概率是多大?

5. 坛子里有 6 个白球 9 个黑球. 如果无放回随机抽取 4 个, 问头两个是白球后两个是黑球的概率是多大?

6. 坛子里有 12 个球, 其中 8 个白球. 从中有放回 (无放回) 抽取 4 个, 求已知抽取的球中正好有 3 个白球的条件下, 第一个球和第三个球是白球的条件概率 (有放回和无放回情形下分别计算).
7. 国王来自有两个孩子的家庭, 问另一个孩子是他姐妹的概率是多大?
8. 某夫妇有两个孩子, 已知老大是女孩的条件下两个孩子都是女孩的条件概率是多大?
9. 假设有 3 个坛子, 坛子 A 有 2 个白球 4 个红球, 坛子 B 有 8 个白球 4 个红球, 坛子 C 有 1 个白球 3 个红球, 如果从每个坛子各取一球, 问正好取了两个白球的条件下, 从坛子 A 里取的是白球的条件概率是多大?
10. 随机无放回地从一副 52 张牌里面抽取 3 张, 已知第二张和第三张都是黑桃, 求第一张是黑桃的条件概率.
11. 随机无放回地从一副 52 张牌中抽取 2 张, 令 B 表示"两张都是 'A' ", 令 A_s 表示"抽中了黑桃 'A' ", 令 A 表示"至少抽中了一张 'A' ".

    (a) 求 $P(B|A_s)$; (b) 求 $P(B|A)$.
12. 某大学毕业生欲参加前三场精算师考试. 她将在 6 月份参加第一场考试. 若通过了, 则在 7 月份参加第二场. 而若又通过了, 则参加 8 月份的第三场. 如果在某场考试失败了, 则不允许参加剩下的考试. 她通过首场考试的概率为 0.9; 如果她通过了首场考试, 则通过第二场考试的条件概率为 0.8; 如果通过了前两场, 那么通过第三场的条件概率为 0.7.

    (a) 她通过全部三场考试的概率是多大?

    (b) 已知她没有通过全部三场考试的条件下, 她在第二场考试失败的条件概率是多大?
13. 考虑一副 52 张牌 (有 4 张"A") 随机平均分给 4 家, 每家 13 张, 我们感兴趣于每家都有一张"A"的概率 p, 令 E_i 表示"第 i 家恰好有 1 张 'A' ", 利用乘法规则计算 $p = P(E_1E_2E_3E_4)$.
14. 坛子里最初有 5 个白球 7 个黑球. 每次取出一个球, 记下它的颜色并放回坛子, 同时再放入同颜色的 2 个球. 计算如下概率:

    (a) 前两个球是黑色, 接下来两个球是白色; (b) 抽取的前四个球当中, 正好有 2 个黑球.
15. 吸烟的怀孕妇女宫外孕的概率是不吸烟妇女的两倍. 如果有 32% 的孕妇吸烟, 那么宫外孕孕妇吸烟的概率是多大?
16. 数据显示, 98% 的婴儿分娩是安全的. 然而有 15% 的分娩是剖腹产. 当采用剖腹产时, 婴儿的生存概率为 96%. 问采用非剖腹产的孕妇, 其婴儿的生存概率是多大?
17. 某个社区, 36% 的家庭有一条狗, 22% 的家庭既有一条狗, 又有一只猫, 另外, 30% 的家庭有一只猫. 求:

    (a) 随机选择一个家庭, 为既有猫又有狗的概率;

    (b) 随机选择一个家庭, 已知该家庭有猫的条件下, 还有一条狗的条件概率.
18. 某城市中, 46% 的人无党派, 30% 的人属于自由党, 24% 的人属于保守党. 在最近一次地方选举中, 35% 的无党派人士、62% 的自由党员、58% 的保守党员参与了选举. 随机选择一位选民, 假定他参与了地方选举, 求以下概率:

    (a) 他是无党派人士; (b) 他是自由党党员;

    (c) 他是保守党党员? (d) 有多少比例的人参与了地方选举?

19. 参加过一个“戒烟班”的人, 有 48% 的女性和 37% 的男性在结束后一年内坚持没有吸烟, 这些人参加了年末的庆功会. 如果一开始, 班里有 62% 的男性, 问:

(a) 参加庆功会的女性有多大比例?

(b) 参加庆功会的人数占全班的百分比是多少?

20. 某大学里, 52% 的学生为女生, 5% 的学生专业为计算机科学, 2% 的学生为计算机科学专业的女生. 如果随机挑选一名学生, 求以下条件概率:

(a) 在已知该学生主修计算机科学的条件下, 该生为女生的条件概率;

(b) 已知该生为女生的条件下, 该生主修计算机科学的条件概率.

21. 总共有 500 对职业夫妇参与了关于年薪的调查, 以下是调查结果:

妻　子	丈　夫	
	低于 25 000 美元	高于 25 000 美元
低于 25 000 美元	212	198
高于 25 000 美元	36	54

举例说明, 其中有 36 对夫妇, 妻子年薪超过 25 000 美元, 而丈夫低于 25 000 美元. 如果随机挑选一对夫妇, 求以下概率:

(a) 丈夫年薪低于 25 000 美元;

(b) 丈夫年薪高于 25 000 美元的条件下, 妻子年薪超过 25 000 美元的条件概率;

(c) 丈夫年薪低于 25 000 美元的条件下, 妻子年薪超过 25 000 美元的条件概率.

22. 分别掷一枚红、蓝、黄色骰子 (都是 6 面), 我们感兴趣于蓝骰子点数小于黄骰子点数, 而黄骰子点数小于红骰子点数的概率. 也即, 令 B, Y, R 分别表示蓝、黄、红骰子的点数, 我们感兴趣 $P(B < Y < R)$.

(a) 没有两个骰子点数一样的概率是多大?

(b) 已知没有两个骰子点数一样的情况下, $B < Y < R$ 的概率是多大?

(c) 求 $P(B < Y < R)$.

23. 坛子 I 有 2 个白球 4 个红球, 而坛子 II 有 1 个白球 1 个红球, 随机从坛子 I 里取一个球放入坛子 II, 然后随机从坛子 II 中取一个球,

(a) 从坛子 II 中取出的球是白球的概率是多大?

(b) 已知从坛子 II 中取出的是白球, 问从第一个坛子中取出的放入第 II 个坛子的球是白球的条件概率是多大?

24. 在一个坛子里放入两个球, 假设在放入之前, 每个球分别以概率为 1/2 涂成黑色, 以概率为 1/2 涂成金色. 假设两个球的涂色是相互独立的.

(a) 假设你已知金色的颜料已经用过 (也即至少有一个球涂成了金色), 计算两个球都涂成金色的概率.

(b) 假设坛子倒了, 一个球掉了出来, 是金色, 那么其中两个球都是金色的概率是多大? 并解释.

25. 以下方法用来估计 100 000 人的城镇里的 50 岁以上的人口的数量: “当你在街上散步时, 数一数你碰到的超过 50 岁的人数, 再算出它们占你遇到的人的百分数, 这样做几天后, 用 100 000 去乘得到的百分数就是所求的估值, ” 对这个方法作出你的评论.

提示: 设这个城市中超过 50 岁的人所占比例为 p, 另外, 令 α_1 表示一个 50 岁以下的人花费在街上的时间所占的比例, α_2 表示一个超过 50 岁的人的相应比值, 这个方法估计的是什么量? 什么时候这个估计值近似等于 p?

26. 假设有 5% 的男性和 0.25% 的女性为色盲, 并假定男性和女性的数量相等. 随机选择一个色盲的人, 他是男性的概率是多大? 如果男性的数量是女性的两倍呢?

27. 一公司所有员工都开车去上班. 公司希望估计出每个车内乘员的平均数. 下面提供的方法中, 哪一个是正确的? 并给出解释.

(a) 随机地找 n 个人, 问他们所乘的车内有多少人, 求出其平均值;

(b) 随机地选 n 辆车, 数一数车内的人数, 然后求平均值.

28. 一副 52 张牌扣在桌上, 每次翻开一张, 直到出现第一张"A". 已知第一张"A"出现在第 20 张翻牌, 问接下来的牌是以下牌的条件概率是多大?

(a) 黑桃"A". (b) 梅花 2.

29. 盒子里有 15 个网球, 其中 9 个球还没用过. 随机抽取 3 个, 用它们练球, 之后放回盒子. 随后, 又随机从中再抽取 3 个, 求其中没有一个球被用过的概率是多大?

30. 两个盒子, 一个里面有黑白弹子各一个, 另一个有 2 个黑弹子, 1 个白弹子. 随机挑选一个盒子, 再随机从中取出一个弹子, 问是黑弹子的概率是多大? 已知弹子是白的条件下, 选中的盒子是第一个盒子的条件概率是多大?

31. 阿奎娜夫人接受了一次癌症活体组织检查. 她不愿意这次检查影响自己周末的情绪. 若她告诉医生, 只有好消息时才打电话通知结果. 这样, 当医生不打电话时, 她仍然可下结论: 她的结果是不好的. 学过概率的阿奎娜夫人要求医生, 首先掷一枚硬币, 若硬币为正面朝上, 那么当检验结果是好消息时, 医生就及时通知她, 当检验结果是坏消息时, 就不通知她. 若硬币为反面朝上时, 医生就不必打电话. 这样, 即使医生不打电话, 并不意味着"一定是坏消息". 记 α 为检查结果为癌症的概率, β 为医生不打电话的条件下, 检验结果为癌症的条件概率.

(a) α 和 β 哪个大? (b) 求出 β 和 α 之间的关系, 验证 (a) 中结论.

32. 某家庭有 j 个孩子的概率为 p_j, 其中 $p_1=0.1, p_2=0.25, p_3=0.35, p_4=0.3$. 随机从该家庭挑选一个孩子, 已知这孩子为该家庭里最大的孩子, 求该家庭有以下数量孩子的条件概率:

(a) 仅仅一个孩子; (b) 有 4 个孩子.

如果随机挑选的一个孩子是该家庭里最小的孩子, 重做 (a) 和 (b).

33. 在下雨天, 乔以概率 0.3 迟到, 而晴天, 乔迟到的概率为 0.1. 根据天气预报, 明天下雨的概率为 0.7.

(a) 求出乔明天按时上班的概率.

(b) 实际上乔在第二天是准时上班的, 问第二天下雨的概率是多少?

34. 在例 3f 中, 假定新的证据仅仅表明罪犯具有该特征的可能性为 90%. 这种情形下, 嫌疑犯确实犯罪的可能性是多大 (假设嫌疑人具有该特征)?

35. 圣诞节送给孩子的礼物都是被藏起来的. 通常以 0.6 的概率是由母亲藏的, 以 0.4 的概率是由父亲藏的. 当母亲藏的时候, 以 70% 的概率藏于楼上, 30% 的概率藏于楼下. 而父亲藏的时候, 则以相等的可能性藏于楼上或楼下.

(a) 礼物放在楼上的概率有多少?

(b) 已知礼物放在楼下, 这个礼物是由父亲藏的概率有多大?

36. 商店 A,B,C 各有 50、75 和 100 名员工, 其中 50%, 60% 和 70% 是女性. 我们假定每个员工的辞职是等可能的, 而且不分员工的性别. 有个员工辞职了, 而且是女性, 问她在 C 店工作的概率是多大?

37. (a) 某赌徒口袋里有一枚均匀硬币, 还有一枚两面都为正面的硬币. 他随机从中取出一枚并掷之, 发现是正面朝上, 问掷的是对称硬币的概率是多大?

(b) 假设他将那枚硬币再掷一次, 发现还是正面朝上. 那么它是对称硬币的概率是多大?

(c) 假设他第三次掷那枚硬币, 发现是反面朝上. 那么它是对称硬币的概率是多大?

38. 坛子 A 里有 5 个白球 7 个黑球. 坛子 B 里有 3 个白球 12 个黑球. 我们投掷一枚对称硬币, 如果结果为正面朝上, 从 A 里取一个球, 而如果结果为反面朝上, 则从 B 里取一个球. 假设取出的是一白球, 问硬币是反面朝上的概率是多大?

39. 在例 3a 中, 已知投保人第一年内没发生事故, 问第二年发生事故的条件概率是多大?

40. 坛子里有 5 个白球, 7 个红球. 考虑如下方式取出 3 个: 每一步取出一个, 并记下颜色, 然后把它放回坛子, 同时再放进一个同颜色的球. 求取出的 3 个球满足下列条件的概率:

(a) 没有白球; (b) 只有 1 个白球; (c) 是 3 个白球; (d) 刚好有 2 个白球.

41. 一副洗好的牌分成两半, 各 26 张. 从其中一半里取出 1 张, 发现是"A". 然后将它放入另一半里, 并将这 27 张牌重新洗匀, 从中取出一张. 计算这张牌是"A"的概率.

提示: 以取出的是否是放进去的那一张为条件.

42. A,B,C 三个厨师烘烤一种饼, 烤坏的概率分别为 0.02, 0.03 和 0.05. 在它们工作的餐厅内, 这种饼, A 烘的占 50%, B 烘的占 30%, C 烘的占 20%. 问在烤坏的饼中, 由 A 烘烤的占多大比例?

43. 盒子里有 3 枚硬币, 第一枚两面都为正面, 第二枚为正常硬币, 第三枚是不对称的, 它出现正面朝上的概率为 75%. 从中随机挑一枚硬币并掷出去, 发现是正面朝上, 问它是两面都为正面的概率是多大?

44. 监狱看守通知三个囚犯, 在他们中要随机选择一个处决, 而把另两个释放. 囚犯 A 请求看守秘密地告诉他, 另外两个囚犯中谁将获得自由, A 声言:"因为我已经知道他们两人中至少有一个获得自由, 所以你泄漏这点消息是无妨的." 但是看守拒绝回答这个问题, 他对 A 说:"如果你知道了你的同伙中谁将获释, 那么, 你自己被处决的概率将由 1/3 增加到 1/2, 因为你就成了剩下的两个未决定命运的囚犯中的一个了." 对于看守的上述理由, 你怎么评价?

45. 假定有 10 枚硬币, 掷第 i 枚硬币正面朝上的概率为 $i/10, i = 1, 2, \cdots, 10$. 先随机选择一枚硬币掷出, 结果为正面, 问它是第 5 枚硬币的概率是多大?

46. 某年内, 投保的男司机索赔的概率为 p_m, 而投保的女司机索赔的概率为 p_f, 其中 $p_f \neq p_m$. 男司机占的比例为 $\alpha, 0 < \alpha < 1$. 随机挑选一名司机, 令 A_i 表示"该司机第 i 年索赔"这一事件, 证明

$$P(A_2|A_1) > P(A_1)$$

给出上述不等式的直观解释.

47. 坛子里有 5 个白球, 10 个黑球. 掷一枚对称骰子, 掷出几点就从坛子中取几个球, 取出的球都是白球的概率是多大? 在取出的球都是白球的条件下, 掷出的骰子点数为 3 的条件概率是多大?

48. 两个外形一样的橱柜都有 2 个抽屉. A 橱柜每个抽屉里有一个银币, B 橱柜有一个抽屉里有一枚银币, 另一抽屉有枚金币. 随机挑选一个橱柜, 打开其中一个抽屉, 发现是一枚银币, 求另一个抽屉里也是银币的概率.

49. 前列腺癌是男性中比较常见的一种癌. 男性是否患有前列腺癌, 医生经常进行一项检查, 测量仅由前列腺分泌的 PSA(prostate specific antigen) 蛋白质水平. 这种检查是出了名地靠不住. 事实上, 一个未患前列腺癌的男性其 PSA 水平偏高的概率为 0.135, 而他确实有癌症的情况下, 此概率增至 0.268. 如果一个医生用其他方法诊断该男士有 70% 的可能患有前列腺癌, 给定下列条件下, 他患有癌症的条件概率是多大?

(a) 检查指标为高水平. (b) 检查指标不为高水平.

若假设医生最初有 30% 的把握认为他有癌症, 重做以上问题.

50. 某保险公司把被保险人分成如下三类:“好的”、“一般的”、“高风险的”. 统计资料表明, 对于上述三种人而言, 在一年期内卷入某一次事故的概率依次为 0.05, 0.15 和 0.30. 如果“好的”被保险人占人口的 20%, “一般的”占 50%, “高风险的”占 30%. 试问在固定的一年中出事故的人口占多大比例? 如果某被保险人在 1997 年没出事故, 他是“好的”(“一般的”) 概率是多大?

51. 某员工为了找到新工作, 请其领导写封推荐信. 她估计如果得到了强有力的推荐, 那么有 80% 的可能找到工作; 如果得到了一般的推荐, 那么 40% 的可能找到工作; 如果得到比较弱的推荐, 那么只有 10% 的可能找到工作. 而且她估计她得到强有力的推荐、一般的推荐和较弱的推荐的概率分别为 0.7,0.2,0.1.

(a) 她有多大的把握认为她会找到新工作?

(b) 已知她找到了新工作, 她得到了强有力的推荐、一般的推荐以及较弱的推荐的条件概率各是多少?

(c) 已知她没找到新工作, 她得到了强有力的推荐、一般的推荐以及较弱的推荐的条件概率各是多少?

52. 一个高中学生非常焦急地等待大学录取通知书. 她估计在她被录取 (不被录取) 的条件下, 在下周各天内收到信件的概率如下表.

日期	P(收到信件 | 被录取)	P(收到信件 | 未录取)
周一	0.15	0.05
周二	0.20	0.10
周三	0.25	0.10
周四	0.15	0.15
周五	0.10	0.20

同时, 估计被录取的概率为 0.6.

(a) 星期一收到信件的概率是多大?

(b) 星期一没收到信件的条件下, 星期二收到信件的概率?

(c) 若星期三以前未收到信件, 她被录取的概率有多大?

(d) 如果她在星期四收到信件, 她被录取的条件概率有多大?

(e) 本周她没收到信件, 问她被录取的条件概率有多大?

53. 并联系统当其中只要有一个元件有效时就工作正常. 考虑一个有 n 个元件的并联系统, 假设每个元件独立地有效工作的概率为 1/2, 计算已知系统工作正常的条件下, 元件 1 工作正常的条件概率.

54. 在下列 (a) 到 (e) 的情形中, 如果你必须建立一个关于事件 E 和 F 的数学模型, 哪种情形你可认为它们是独立的? 试解释原因.

(a) E 表示某个女商人是蓝眼睛, 而 F 表示她的秘书也是蓝眼睛.

(b) E 表示某教授有辆汽车, F 表示他的名字出现在电话簿里.

(c) E 表示某男人身高低于 6 英尺, 而 F 表示其体重超过 200 磅.

(d) E 表示某妇女生活在美国, 而 F 表示她生活在西半球.

(e) E 表示明天会下雨, 而 F 表示后天还会下雨.

55. 某课堂里已知有 4 个一年级男生, 6 个一年级女生, 还有 6 个二年级男生. 如果从中随机选择一个学生, 且其性别和年级是独立的, 那么此课堂有多少个二年级女生?

56. 你一直在收集优惠券, 假设一共有 m 种优惠券. 若每次收集优惠券时, 得到第 i 种优惠券的概率为 $p_i, i = 1, \cdots, m$. 假设你正收集第 n 张优惠券, 那么它是一张新类型 (以前未曾收集到) 的概率有多大?

提示: 以这张优惠券的种类为条件.

57. 一个关于股价变化的简化模型为: 每一天股价上涨一个单位的概率为 p, 下跌一个单位的概率为 $1-p$, 不同天里的变化认为是独立的.

(a) 2 天后股价仍维持在初始水平的概率是多大?

(b) 3 天后股价上涨 1 个单位的概率是多大?

(c) 已知三天后股价上涨了一个单位, 问第一天股价上涨的条件概率是多大?

58. 我们希望模拟掷一枚均匀硬币的试验, 但是我们只有一枚不一定均匀的硬币, 其正面朝上的概率为未知的 p(p 不一定为 1/2). 考虑下面步骤:

(1) 掷一次硬币; (2) 再掷一次硬币;

(3) 如果上面两次结果一样, 回到第 1 步; (4) 最后一次掷出的结果, 作为试验结果.

(a) 证明: 这样得到的结果, 正面朝上或朝下的概率是一样的.

(b) 如果我们简化了步骤. 将硬币掷到出现两次不同时为止, 将最后一次掷硬币的结果作为试验结果. 这样做可以吗?

59. 独立抛掷一枚硬币, 每次正面朝上的概率为 p, 前四次结果如下的概率是多大?

(a) H,H,H,H; (b) T,H,H,H;

(c) 在连续掷硬币过程中, T,H,H,H 出现在 H,H,H,H 之前的概率是多大?

提示: 对于 (c), 在什么条件下 H,H,H,H 先发生?

60. 人眼的颜色由一对基因决定. 如果都为蓝色基因, 那么眼睛为蓝色; 如果都为棕色基因, 那么眼睛为棕色. 如果一个是蓝色基因, 一个是棕色基因, 那么眼睛为棕色. (我们称棕色基因比蓝色基因占优势.) 一个新生儿独立地从其父母处各得到一个遗传基因, 父亲的基因对中的任一基因以相等的概率遗传给他的孩子. 母亲的情况也是一样的. 假设史密斯及其父母眼睛都为棕色, 但其姐姐眼睛为蓝色.

(a) 史密斯拥有蓝色基因的概率多大? 假设史密斯先生的夫人眼睛为蓝色.

(b) 他们第一个孩子的眼睛为蓝色的概率是多大?

(c) 如果他们第一个孩子的眼睛为棕色, 它们第二个孩子眼睛为棕色的概率是多大?

61. 有关白化病的基因记为 A 和 a, 只有从其父母都遗传得到了基因 a 的人才会得白化病. 有基因对 A,a 的人在外表上是正常的, 但是因为他能将其基因传给下一代, 因此称为携带者. 现假设有一对正常的夫妇有两个小孩, 其中有一个为白化病. 假设另一个未得白化病的孩子将来与一个白化病携带者结婚.

(a) 他们的第一个孩子得白化病的概率是多大?

(b) 已知他们的第一个孩子未得白化病的条件下, 第二个孩子得白化病的条件概率是多大?

62. 芭芭拉和黛安娜出去射击. 假设芭芭拉每次射击击中目标 (木鸭子) 的概率为 p_1, 而黛安娜每次射击击中目标的概率为 p_2. 假设她们同时射击同一目标. 如果木鸭子翻了 (表示射中了), 以下事件概率是多大?

(a) 两人都射中了. (b) 芭芭拉射中了.

其中作了怎样的独立性假设?

63. A 和 B 卷入一场决斗. 决斗的规则是: 捡起自己的枪, 并同时射击对方. 如果有一人或两人都被射中, 那么决斗结束. 如果两人都射空, 那么重复过程. 假设每次射击结果都是独立的, 且 A 射中 B 的概率为 p_{A}, B 射中 A 的概率 p_{B}, 计算

(a) A 没被击中的概率; (b) A,B 都被击中的概率;

(c) n 局决斗后决斗停止的概率; (d) A 没被击中的条件下, n 局决斗停止的条件概率;

(e) 两人都被击中的条件下, n 局决斗停止的条件概率.

64. 在一个答题秀节目上, 一对夫妇遇到了一道题, 丈夫和妻子独立给出正确答案的概率都为 p. 对于这对夫妇, 以下哪个策略更好?

(a) 任选一个人并让其答题.

(b) 他们俩都给出问题的答案, 如果答案一致, 那么就采用这答案, 如果答案不一致, 那么掷硬币决定采取谁的答案.

65. 在 习题 64 中, 若 $p = 0.6$, 且夫妇采用 (b) 的策略, 那么在如下条件下夫妇给出正确答案的条件概率是多大?

(a) 夫妇给出的答案一致. (b) 夫妇给出的答案不一致.

66. 在图 3.4 中所示的电路里, 第 i 个继电器闭合的概率为 $p_i, i = 1,2,3,4,5$. 如果所有继电器的功能相互独立, 试对如下 (a)、(b) 两种情况, 分别求出 A 和 B 之间是通路的概率.
提示: 对于 (b), 以继电器 3 是否闭合为条件.

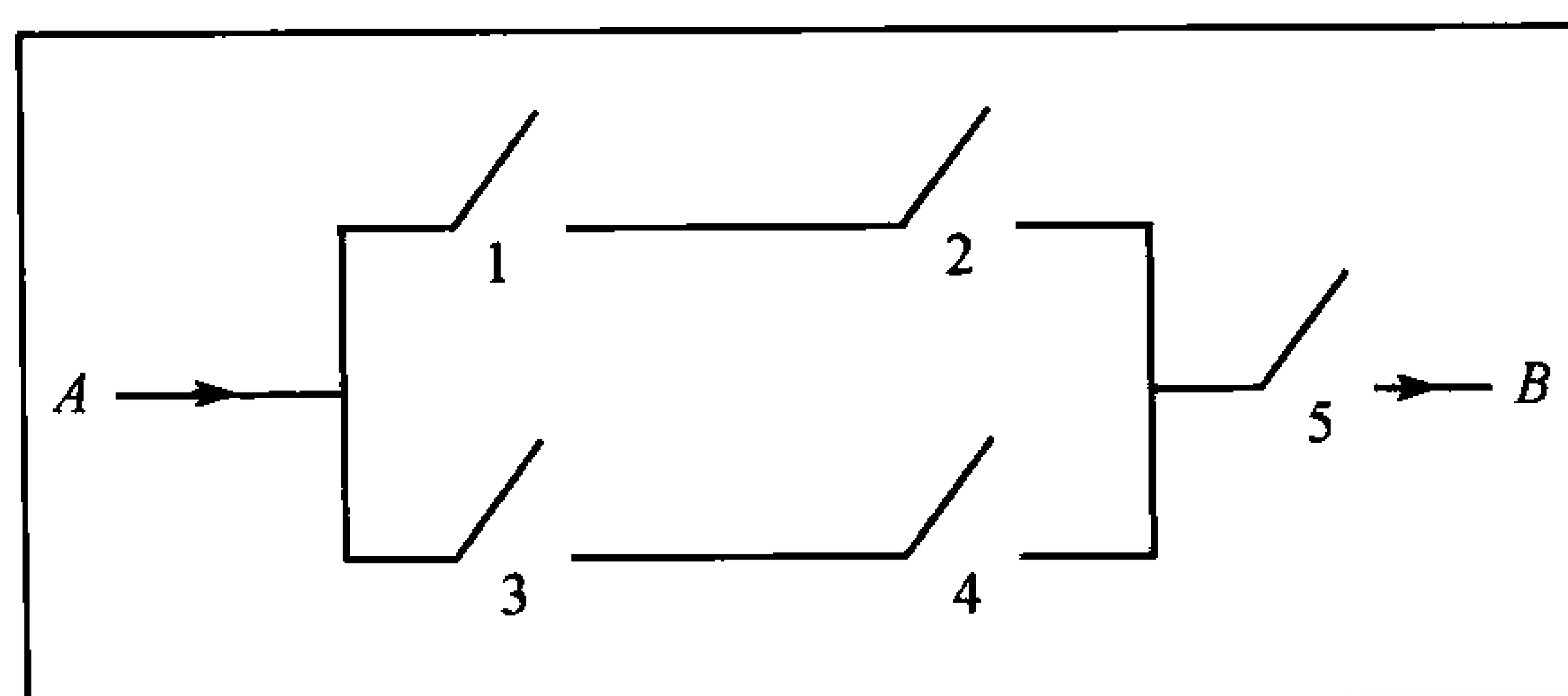
67. 一个由 n 个部件组成的系统称为"$k - n$"系统 ($k \leqslant n$), 如果此系统运行当且仅当它的 n 个部件中至少有 k 个部件运行. 假设它的各部件运行是相互独立的.

(a) 如果第 i 个部件运行的概率为 $p_i, i = 1,2,3,4$, 求一个"2 - 4"系统运行的概率.

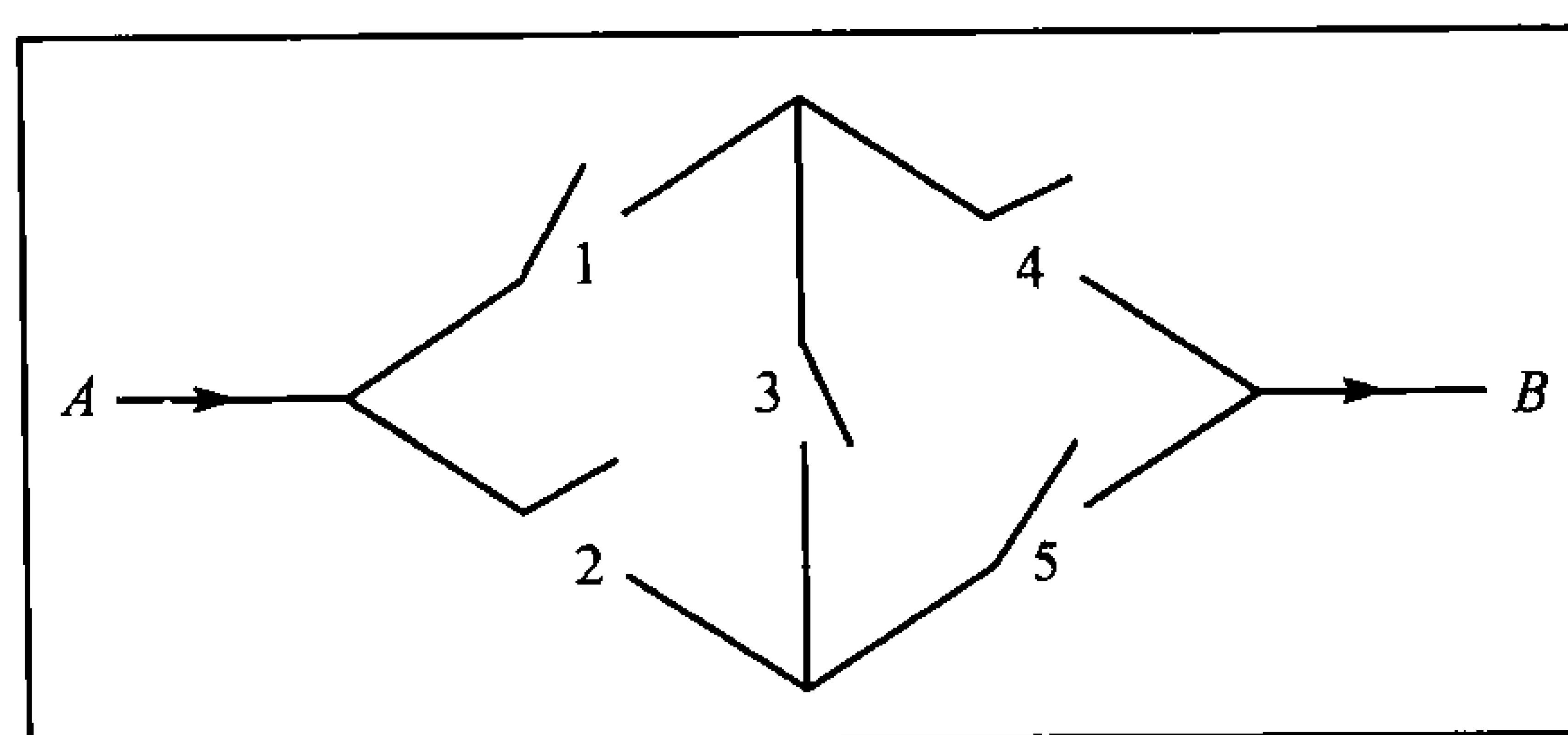
(b) 试对"3 - 5"系统求出上述概率;

(c) 假设一个"$k - n$"系统所有的 p_i 都等于 p(即 $p_i = p, i = 1,2\cdots,n$), 试对此系统求出上述概率.

68. 在 习题 66(a) 中, 已知 A 和 B 是通路的条件下, 求继电器 1 和 2 均闭合的概率.



(a)



(b)

图 3.4　习题 66 中的电路示意图

69. 某种生物具有一对基因组, 其中每个基因组由 5 个不同的基因组成. 这 5 个基因用 5 个英文字母表示. 每个基因有两种特性, 分别用大小字母来区别. 大写字母表示显性, 小写字母表示隐性. 若某生物具有基因对 (x, X), 此时, 该生物表现基因 X 的特征. 例如, X 代表棕色眼睛, x 代表蓝色眼睛, 那么具有 (X,X) 或 (x,X) 的都是棕色眼睛, 而 (x,x) 时为蓝色眼睛. 生物的表现特征 (phenotype) 与基因特征 (genotype) 是有区别的. 例如, 一个生物具有基因特征 aA,bB,cc,dD,ee, 另一个生物具有基因特征 AA,BB,cc,DD,ee, 它们的基因特征不同, 但表现特征是相同的. 对于每一种基因的基因对, 交配双方都随机地贡献其中一个. 例如, 对某种基因, 父方的基因为 (X,x), 母方的基因为 (X,X), 此时, 其子女接受父方的一个 X(或 x), 母方的一个 X, 构成子女的基因的两个特征. 现在设交配双方的 5 种基因组分别为 aA,bB,cC,dD,eE 和 aa,bB,cc,Dd,ee. 分别计算子女的 (i) 基因特征和 (ii) 表现特征与 (a) 第一个亲代相同的概率; (b) 第二个亲代相同的概率; (c) 某一个亲代相同的概率; (d) 与两个亲代都不相同的概率.

(第一个亲代指基因组为 aA, bB, cC, dD, eE 的亲代, 第二个亲代指基因组为 aa, bB, cc, Dd, ee 的亲代.)

70. 女王有 50% 的可能携带有血友病的基因. 如果她是一个携带者, 那么每个王子都有 50% 的可能有血友病. 如果女王有 3 个王子, 且都没有血友病. 那么女王是携带者的概率有多大? 如果有第四个王子, 那么他有血友病的概率有多大?

71. 在 1982 年 12 月 30 日, 美国全国棒球协会西部赛区积分榜排名前三名如下:

球 队	赢	输
亚特兰大勇士	87	72
旧金山巨人	86	73
洛杉矶躲闪者	86	73

每个队还剩三场比赛需要打, 巨人队的所有三场比赛的对手是躲闪者, 而勇士队的所有三场比赛的对手是圣迭戈教士队. 假设所有比赛结果是独立的, 且每场比赛双方获得胜利的可能性是一样的. 那么各个队取得排名第一的概率是多大? (如果两个队并列第一, 还需要进行附加赛, 以决定名次, 此时, 假定各队获胜概率为 1/2.)

72. 市政委员会由 7 人组成, 其中包含一个由 3 人组成的核心委员会. 一个新的法律提案首先由核心会员会表决. 若核心委员会中有 2 人以上同意, 才能拿到全体委员会上讨论. 一旦到了全体委员会, 只要有 4 票以上通过, 这个新的法律提案就生效. 现在有一新的法律提案. 设每个委员以概率 p 同意这个提案, 并且相互独立地作出投票决定. 求一个核心委员会成员起决定作用的概率. 所谓起决定作用是指若他的决定是相反, 其最后结果也相反. 即若他投赞成票, 法案就通过, 他若投反对票, 法案就不通过. 不在核心委员会的委员起决定作用的概率是多少?

73. 假设某对夫妇的每个孩子是男是女的可能性一样, 且与这个家庭里的其他孩子的性别是独立的. 对于一个有 5 个孩子的家庭, 计算如下事件的概率:

(a) 所有的小孩同一性别; (b) 最大的三个为男孩, 其他为女孩;

(c) 正好三个男孩; (d) 最大的两个为女孩; (e) 至少有一个女孩.

74. A 和 B 轮流掷一对骰子, 当 A 掷出 "和为 9" 或 B 掷出 "和为 6" 时停止, 假设 A 先掷, 求最后一次掷是由 A 完成的概率.

75. 某村子有一个传统, 家庭里的长子和其妻子有义务照顾他们的年迈父母. 然而, 近年来, 这个村的妇女不愿意嫁给长子.

(a) 如果每个家庭都有两个孩子, 问所有的儿子中长子占多大比例?

(b) 如果每个家庭都有三个孩子, 问所有的儿子中长子占多大比例?

假设出生婴儿时, 是男孩或女孩的概率是相同的, 并且家庭里各小孩的性别是独立的.

76. 假设 E, F 为某次试验的互不相容的事件. 证明: 如果独立重复进行这样的试验, 那么 E 发生在 F 之前的概率为 $P(E)/[P(E)+P(F)]$.

77. 考虑一个无穷的独立重复试验序列, 每次试验都等可能地以 1, 2, 3 为结果. 假设 3 次试验后最后一次试验结果为 3, 求以下条件概率:

(a) 第一次试验结果为 1; (b) 前两次试验结果都为 1.

78. A 和 B 进行一系列比赛, 每局比赛 A 获胜的概率都为 p, B 获胜的概率为 $1-p$, 且每次比赛结果相互独立. 当其中一人比另一人多胜两局时, 游戏停止, 获胜局数多的选手获得比赛胜利.

(a) 求总共比赛了 4 局的概率. (b) 求 A 最后获得比赛胜利的概率.

79. 连续地掷一对均匀骰子, 在出现 6 次 "和为偶数" 之前, 出现 2 次 "和为 7" 的概率是多大?

80. 选手们水平相当, 在每次比赛中任一方获胜的可能性都是 1/2. 共有 2^n 个选手随机地一一配对比赛, 随后的 2^{n-1} 个胜者再随机地一一配对比赛, 如此这般, 直到最后一个获胜者出现. 对指定的 A, B 两人, 定义事件 $A_i, i \leqslant n$ 和 E 如下:

$$A_i: \text{A 参与了 } i \text{ 场比赛} \quad E: \text{A,B 从未比赛过}$$

(a) 求 $P(A_i), i=1,\cdots,n$;　(b) 求 $P(E)$.

(c) 令 $P_n = P(E)$, 证明

$$P_n = \frac{1}{2^n-1} + \frac{2^n-2}{2^n-1}\left(\frac{1}{2}\right)^2 P_{n-1}$$

并利用此验证 (b) 里得到的答案.

提示: 以事件 $A_i, i=1,\cdots,n$ 发生为条件, 计算 $P(E)$, 再利用以下代数恒等式

$$\sum_{i=1}^{n-1} ix^{i-1} = \frac{1-nx^{n-1}+(n-1)x^n}{(1-x)^2}$$

来化简答案. 另外一个解决此问题的方法是, 注意到一共进行了 2^n-1 场比赛.

(d) 解释为什么总共有 2^n-1 场比赛.

给这些比赛进行编号, 且令 B_i 表示“A 和 B 在第 i 场比赛中碰面”, $i=1,\cdots,2^n-1$.

(e) $P(B_i)$ 是多少?　(f) 利用 (e) 计算 $P(E)$.

81. 某股票投资者有一只股票, 其现价为 25. 她决定当股价跌至 10 或涨至 40 时卖出股票. 如果股价的每次变化都是以 0.55 的概率上涨 1 点, 以 0.45 的概率下跌 1 点, 且每次变动 (上涨或下跌) 是相互独立的, 那么投资者最后获利的可能性是多大?

82. A 和 B 掷硬币, A 先开始并连续掷硬币, 直到出现反面朝上. 此时, B 开始掷硬币直到出现反面朝上, 然后又轮到 A 掷, 等等. 令 P_1, P_2 分别表示 A 和 B 掷硬币时正面朝上的概率. 若谁先达到如下获胜条件则胜, 求 A 获胜的概率.

(a) 在一轮中有两个正面朝上;　(b) 双方总共 2 个正面朝上;

(c) 一轮中 3 个正面朝上;　(d) 双方总共 3 个正面朝上.

83. 骰子 A 有 4 面红 2 面白, 而骰子 B 有 2 面红 4 面白. 先掷一枚均匀的硬币, 如果正面朝上, 那么用骰子 A 玩游戏, 如果是反面朝上, 那么用骰子 B. (在整个游戏过程中, 只在开始时掷一次硬币.)

(a) 证明每次掷出红色一面的概率为 1/2;

(b) 若前两次投掷结果都为红色一面, 那么第三次投掷也为红色的概率是多大?

(c) 如果头两次投掷都为红色, 那么使用的是骰子 A 的概率是多大?

84. 坛子里有 12 个球, 其中 4 个白球, 三个选手 A,B,C 依次从坛中取球, A 最先, 然后 B, 然后 C, 再是 A, 依次进行下去.第一个取出白球的人获胜. 求每个选手获胜的概率, 如果

(a) 每个球取出后再放回;　(b) 取出的球不放回.

85. 当 3 个选手各自选择自己的坛子时, 重做习题 84, 也即, 假设有 3 个不同的坛子, 每个里面有 12 个球, 其中 4 个白球.

86. 令 $S=\{1,2,\cdots,n\}$, 假设 A 和 B 独立, 等可能地为 2^n 个子集之一 (包括空集和 S 本身).

(a) 证明:

$$P\{A \subset B\} = \left(\frac{3}{4}\right)^n$$

提示: 令 $N(B)$ 表示 B 里元素个数, 利用

$$P\{A \subset B\} = \sum_{i=0}^{n} P\{A \subset B | N(B) = i\} P\{N(B) = i\}$$

(b) 证明 $P\{AB = \varnothing\} = \left(\frac{3}{4}\right)^n$.

87. 在例 5e 里, 已知前 n 次结果都是正面朝上的条件下, 求选中第 i 个硬币的条件概率.

88. 拉普拉斯继承规则里 (例 5e), 各次投掷结果是否独立? 试解释之.

89. 由 3 名法官组成一陪审团. 某人被审讯后, 若至少两名法官投“有罪”票, 则判决此人有罪. 假设对一名事实上有罪的被告, 每个法官独立地投“有罪”票的概率为 0.7, 而对事实上无罪的被告, 这个概率下降到 0.2. 如果 70% 的被告事实上是有罪的, 试对如下各条件求出 3 号法官投“有罪”票的条件概率.

(a) 1 号法官和 2 号法官都投“有罪”票;

(b) 在 1 号和 2 号法官所投的票中, 一张“有罪”, 另一张“无罪”;

(c) 1 号和 2 号法官全投“无罪”票.

以 $E_i, i = 1, 2, 3$ 表示 i 号法官投一张“有罪”票的事件, 这些事件是否相互独立? 是否条件独立? 试说明理由.

90. 假设进行 n 次独立重复试验, 每次结果是 0,1 或 2 的概率分别是 p_0, p_1 和 p_2, $\sum_{i=0}^{2} p_i = 1$. 求结果 1 和 2 都至少出现一次的概率是多大?

理 论 习 题

1. 若 $P(A) > 0$, 证明

$$P(AB|A) \geqslant P(AB|A \cup B)$$

2. 若 $A \subset B$, 尽可能化简如下概率:

$$P(A|B) \qquad P(A|B^c) \qquad P(B|A) \qquad P(B|A^c)$$

3. 考虑有 m 个家庭的社区, 其中有 n_i 个家庭有 i 个孩子 $i = 1, \cdots, k$, $\sum_{i=1}^{k} n_i = m$. 考虑如下两种选择孩子方式:

(a) 随机选择一个家庭, 然后再随机选择一个孩子;

(b) 从 $\sum_{i=1}^{k} i n_i$ 个孩子中随机挑选一个.

证明: 第一种方法挑选出来的孩子是他家里第一个出生的孩子的概率大于第二种方法挑选出来的.

提示: 为了解此题, 需要证明

$$\sum_{i=1}^{k} i n_i \sum_{j=1}^{k} \frac{n_j}{j} \geqslant \sum_{i=1}^{k} n_i \sum_{j=1}^{k} n_j$$

为了证明这点, 将两边展开, 证明对于任意一组 (i, j), 式子左边的项 $n_i n_j$ 的系数比右边的大.

4. 已知有一个球放在 n 个盒子中的一个内. 已知球在第 i 个盒子的概率是 P_i. 如果球在第 i 个盒子里, 搜寻该盒子时还可以找不到这颗球. 假定球在盒子 i 内的条件下, 搜寻时能找到球的概率为 α_i. 证明: 已知搜索第 i 个盒子没有发现球的条件下, 球在第 j 个盒子的条件概率是

$$\begin{cases} \dfrac{P_j}{1-\alpha_i P_i} & 若 j \neq i \\ \dfrac{(1-\alpha_i)P_i}{1-\alpha_i P_i} & 若 j = i \end{cases}$$

5. 事件 F 称作不利于事件 E 的, 并记为 $F \searrow E$, 若

$$P(E|F) \leqslant P(E)$$

试证明如下论断或者给出反例:

(a) 若 $F \searrow E$, 则 $E \searrow F$;

(b) 若 $F \searrow E$, 且 $E \searrow G$, 则 $F \searrow G$;

(c) 若 $F \searrow E$, 且 $G \searrow E$, 则 $FG \searrow E$.

同样还可定义 F 有利于 E, 并记作 $F \nearrow E$, 若 $P(E|F) \geqslant P(E)$. 将 $\searrow$ 改为 $\nearrow$ 重做 (a)(b)(c).

6. 证明: 若 $E_1, E_2, \cdots, E_n$ 为相互独立事件列, 则

$$P(E_1 \cup E_2 \cup \cdots \cup E_n) = 1 - \prod_{i=1}^{n}[1 - P(E_i)]$$

7. (a) 坛子里有 n 个白球和 m 个黑球, 每次随机从中取出一个, 直到剩下的球为同一种颜色. 证明剩下的球全为白球的概率为 $n/(n+m)$;

提示: 设想试验一直进行直到所有球都取走, 并考虑最后取出的球的颜色.

(b) 池塘里有 3 种不同的鱼, 我们分别称为红、蓝、绿鱼. 三种鱼分别有 r、b、g 条. 假设按一随机顺序将这些鱼全部移走 (即每次选择对于剩下的鱼都是等可能的), 问最先抓完红鱼的概率?

提示: 将蓝、绿的鱼视为同一种鱼, 称为非红鱼. 再利用 (a) 中采用的方法解题.

8. 设一次掷两枚对称的骰子, A, B, C 为与试验有关的三个事件.

(a) 若

$$P(A|C) > P(B|C) \quad 且 \quad P(A|C^c) > P(B|C^c)$$

或者证明 $P(A) > P(B)$, 或者定义事件 A, B, C 说明不等式 $P(A) > P(B)$ 不成立.

(b) 若

$$P(A|C) > P(A|C^c) \quad 且 \quad P(B|C) > P(B|C^c)$$

或者证明 $P(AB|C) > P(AB|C^c)$, 或者构造事件 A, B, C 说明不等式 $P(AB|C) > P(AB|C^c)$ 不成立.

提示: 令 C 表示 "掷一对骰子, 点数之和为 10", 令 A 表示 "第一个骰子点数为 6", 令 B 表示 "第二个骰子点数为 6".

9. 考虑独立地掷两次均匀硬币. 令 A 表示第一次为正面朝上, 令 B 表示第二次为正面朝上, 令 C 表示两次朝向一样. 证明事件 A, B, C 为两两独立. 也即, A 和 B 独立, A 和 C 独立, B 和 C 独立, 但 A, B, C 不独立.

10. 考虑有 n 个人, 假定每人的生日在 365 天内任一天的可能性都是一样的, 且相互独立. 令 $A_{i,j}$ 表示"第 i 个人和第 j 个人"生日相同, $i \neq j$, 证明: 这些事件都是两两独立的, 也即 $A_{i,j}$ 和 $A_{r,s}$ 是独立的, 但是这 $\binom{n}{2}$ 个事件 $A_{i,j}, i \neq j$ 并不独立.

11. 设有一枚硬币, 在掷硬币时正面朝上的概率为 p. 现独立掷 n 次. 问 n 需多大, 才能保证至少有一次正面朝上的概率大于 1/2?

12. 若 $0 \leqslant a_i \leqslant 1, i = 1, 2, \cdots$, 证明

$$\sum_{i=1}^{\infty}\left[a_i \prod_{j=1}^{i-1}(1-a_j)\right] + \prod_{i=1}^{\infty}(1-a_i) = 1$$

提示: 假设掷无限多个硬币, 令 a_i 表示第 i 个硬币正面朝上的概率, 考虑出现第一个正面朝上的时刻.

13. 设有一枚硬币, 在抛掷时正面朝上的概率为 p. 假设 A 开始连续掷硬币, 直至出现了反面朝上, 此时 B 接着连续掷硬币, 直到掷出反面朝上为止, 然后 A 再接着掷, 如此进行下去. 令 $P_{n,m}$ 表示在 B 累计 m 次正面朝上之前 A 已累计 n 次正面朝上的概率. 证明:

$$P_{n,m} = pP_{n-1,m} + (1-p)(1-P_{m,n})$$

***14.** 假设你同一个无限富裕的人赌博, 每一步你可能赢也有可能输 1 个单位, 概率分别为 p 和 $1-p$. 证明你最后输光的概率为

$$\begin{cases} 1 & 若 p \leqslant 1/2 \\ (q/p)^i & 若 p > 1/2 \end{cases}$$

其中 $q = 1 - p$ 且 i 是你的最初资金.

15. 独立重复进行每次成功率为 p 的试验, 直到总共有 r 次成功为止. 证明恰好需要进行 n 次试验的概率为

$$\binom{n-1}{r-1} p^r (1-p)^{n-r}$$

并利用此结果解决 赌本分割问题(例 4j).

提示: 为了在 n 次试验中得到 r 次成功, 在前 $n-1$ 次试验中必须有多少次成功?

16. 若某独立重复实验序列中, 每次试验结果只有两种, 成功或失败, 则这种试验序列称为伯努利试验序列. 现设一个 n 次独立重复的伯努利试验序列, 每次试验成功的概率为 p(失败的概率为 $1-p$). 记 P_n 表示 n 试验中出现偶数次 (0 认为是偶数) 成功的概率, 证明

$$P_n = p(1-P_{n-1}) + (1-p)P_{n-1} \qquad n \geqslant 1$$

并利用此公式 (利用归纳法) 证明

$$P_n = \frac{1 + (1-2p)^n}{2}$$

17. 考虑进行 n 次独立试验, 第 i 次试验成功的概率为 $1/(2i+1)$. 令 P_n 表示总的成功次数为奇数的概率.

(a) 计算 P_n, $n=1,2,3,4,5$;

(b) 猜测 P_n 的一个一般公式;

(c) 导出用 P_{n-1} 表示的 P_n 的递推公式;

(d) 验证 (b) 中的猜测满足 (c) 中的递推公式. 因为递推公式只有一个解, 这就证明了你的猜测是正确的.

18. 令 Q_n 表示连续掷 n 次均匀硬币, 没有出现 3 个连续的正面朝上的概率. 证明

$$Q_n = \frac{1}{2}Q_{n-1} + \frac{1}{4}Q_{n-2} + \frac{1}{8}Q_{n-3}$$

$$Q_0 = Q_1 = Q_2 = 1$$

并求 Q_8.

提示: 以出现第一次反面朝上的时刻为条件.

19. 考虑赌徒输光问题, 不同的是 A 和 B 同意最多玩 n 局. 令 $P_{n,i}$ 表示 A 最后输光所有钱的概率, 其中开始时 A 有 i 元而 B 有 $N-i$ 元. 推导用 $P_{n-1,i+1}$ 和 $P_{n-1,i-1}$ 表示的 $P_{n,i}$ 的公式, 并计算 $P_{7,3}$, $N=5$.

20. 假设有两个坛子, 每个坛子中既有白球, 也有黑球. 从两个坛子里取出白球的概率分别为 p 和 p'. 按照如下方式连续有放回地取球. 最初分别以 α 的概率从第一个坛子里取球, 以 $1-\alpha$ 的概率从第二个坛子里取球. 接下来的取球按如下规则进行: 一旦取出的是白球, 则放回, 然后从同一个坛子里再取球; 如取出的是黑球, 那么接下来从另一个坛子里取球. 令 α_n 表示第 n 次从第一个坛子里取球的概率, 证明

$$\alpha_{n+1} = \alpha_n(p+p'-1) + 1 - p' \qquad n \geqslant 1$$

并用此证明

$$\alpha_n = \frac{1-p'}{2-p-p'} + \left(\alpha - \frac{1-p'}{2-p-p'}\right)(p+p'-1)^{n-1}$$

令 P_n 表示第 n 次取出的球是白球的概率, 求 P_n. 并且计算 $\lim\limits_{n\to\infty} \alpha_n$ 和 $\lim\limits_{n\to\infty} P_n$.

21. 投票问题. 在一次选举中, 候选人 A 获得了 n 张选票, 而 B 获得了 m 张选票, 其中 $n>m$. 假定所有的 $(n+m)!/(n!m!)$ 种选票的顺序都是等可能的, 令 $P_{n,m}$ 表示在统计选票时 A 一直领先的概率.

(a) 计算 $P_{2,1}, P_{3,1}, P_{3,2}, P_{4,1}, P_{4,2}, P_{4,3}$.

(b) 求 $P_{n,1}, P_{n,2}$.

(c) 基于 (a) 和 (b) 的结果, 猜测 $P_{n,m}$ 的值.

(d) 以谁获得了最后一张选票为条件, 求用 $P_{n-1,m}$ 和 $P_{n,m-1}$ 表示的 $P_{n,m}$ 的递推公式.

(e) 利用 (d), 对 $n+m$ 用归纳法证明 (c) 中的猜测.

22. 作为天气预报的一个简单模型, 假设明天天气 (湿润或者干燥) 与今天相同的概率为 p. 如果 1 月 1 日天气为干燥, 证明, n 天后的天气为干燥的概率 P_n, 满足

$$P_n = (2p-1)P_{n-1} + (1-p) \qquad n \geqslant 1$$

$$P_0 = 1$$

并证明：

$$P_n = \frac{1}{2} + \frac{1}{2}(2p-1)^n \qquad n \geqslant 0$$

23. 设一个包里有 a 个白球, b 个黑球, 根据以下方式从包里取球.

第 1 步　随机取出一个球, 然后扔掉;

第 2 步　再随机取出一个球, 若取出的球的颜色与前一次取出的不同, 将球放回包内, 转向第 1 步. 否则, 将球扔掉, 重复第 2 步.

也就是说, 随机取球并扔掉, 直至颜色发生了变化, 此时将取到的球放回包里重新开始. 令 $P_{a,b}$ 表示包里最后一个球为白球的概率, 证明：

$$P_{a,b} = \frac{1}{2}$$

提示：对 $k \equiv a+b$ 用归纳法.

***24.** n 个选手的单循环赛, $\binom{n}{2}$ 对选手只比赛一次, 任何一场比赛的结果都分出输赢. 对于一个给定的整数 $k, k<n$, 一个有趣的问题是：是否可能存在一种比赛结果, 对于任意 k 个选手的集合, 都有一位选手, 他打败了该集合内的所有选手. 证明：若

$$\binom{n}{k}\left[1-\left(\frac{1}{2}\right)^k\right]^{n-k} < 1$$

则这种结果是存在的.

提示：假设比赛结果是相互独立的, 且每场比赛谁获胜的可能性也是一样的. 给 k 个选手的所有 $\binom{n}{k}$ 个组合标号, 令 B_i 表示"在第 i 个组合里, 没有人打败所有 k 个选手", 然后利用布尔不等式求出 $P\left(\bigcup_i B_i\right)$ 的界.

25. 直接证明

$$P(E|F) = P(E|FG)P(G|F) + P(E|FG^c)P(G^c|F)$$

26. 证明 (5.11) 和 (5.12) 的等价性.

27. 将条件独立性的定义推广到 2 个以上的事件的情形.

28. 证明或给出反例：如果 E_1 和 E_2 是独立的, 那么给定 F 的条件下, 它们也条件独立.

29. 在拉普拉斯继承准则 (例 5e) 中, 证明：已知前 n 次掷的结果都是正面朝上的条件下, 接下来 m 次掷的结果也是正面朝上的条件概率为 $(n+1)/(n+m+1)$.

30. 在拉普拉斯继承准则中, 假设前 n 次投掷中, 有 r 次正面朝上, $n-r$ 次反面朝上. 证明：第 $n+1$ 次投掷正面朝上的条件概率为 $(r+1)/(n+2)$. 此题目证明必需先证明恒等式

$$\int_0^1 y^n(1-y)^m \, \mathrm{d}y = \frac{n!\,m!}{(n+m+1)!}$$

提示：为了证明该恒等式, 令 $C(n,m) = \int_0^1 y^n(1-y)^m \, \mathrm{d}y$. 通过分部积分可得：

$$C(n,m) = \frac{m}{n+1} C(n+1, m-1)$$

从 $C(n,0) = 1/(n+1)$ 开始, 对 m 利用归纳法证明恒等式.

31. 假定你的一个不懂数学但有哲学头脑的朋友声称拉普拉斯继承准则必定是错误的, 理由是它可以导出荒谬的结论. "例如, "他说, "如果一个孩子 10 岁, 按照这个原则, 由于他已经活了 10 年, 那么再活一年的概率为 11/12. 另一方面, 如果孩子有位 80 岁的祖父, 那么由拉普拉斯准则, 祖父再活一年的概率为 81/82. 显然这是荒谬的, 因为孩子比祖父更容易再活一年. "你怎么回答这位朋友?

自 检 习 题

1. 在一局桥牌中, 西家没有“A”, 求以下事件的概率:

(a) 他的搭档没有“A”; (b) 他的搭档有 2 个或更多的“A”?

(c) 如果西家正好有一个“A”, 以上概率又各是多少?

2. 一个新的汽车电池使用超过 10 000 英里的概率为 0.8, 超过 20 000 英里的概率为 0.4, 超过 30 000 英里的概率为 0.1. 如果一个新的电池在使用 10 000 英里以后仍能继续使用, 求以下概率:

(a) 其总寿命超过 20 000 英里; (b) 它还能继续使用 20 000 英里.

3. 怎样将 10 个白球 10 个黑球共 20 个球放入两个坛子内, 使得随机选择一个坛子并随机从中取出一个球是白球的概率最大?

4. 坛子 A 有 2 个白球 1 个黑球, 而坛子 B 有 1 个白球 5 个黑球. 随机从坛子 A 中取一个球放入坛子 B, 然后再随机从坛子 B 内取出一球. 已知取出的是白球, 问从坛子 A 中取出的放入坛子 B 的球是白球的概率是多大?

5. 坛子里有 r 个红球和 w 个白球, 每次随机取走一个. 令 R_i 表示“第 i 次取走的是红球”, 计算

(a) $P(R_i)$; (b) $P(R_5|R_3)$; (c) $P(R_3|R_5)$.

6. 坛子里有 b 个黑球, r 个红球. 随机取出一个, 再放回, 同时还放入 c 个同样颜色的球. 现在, 再取一个球, 证明已知第二个球是红球的条件下, 第一个球是黑球的概率为 $b/(b+r+c)$.

7. 你的一个朋友从一副 52 张牌里无放回地随机取出 2 张牌. 在下列情形下, 计算两张都是“A”的条件概率:

(a) 你问是否其中一张是黑桃“A”, 你朋友的回答是肯定的;

(b) 你问是否第一张是“A”, 你朋友的回答是肯定的;

(c) 你问是否第二张是“A”, 你朋友的回答是肯定的;

(d) 你问是否其中有一张是“A”, 你朋友的回答是肯定的.

8. 证明:

$$\frac{P(H|E)}{P(G|E)}=\frac{P(H)}{P(G)}\frac{P(E|H)}{P(E|G)}$$

考虑: 在得到新的证据前, 假设 H 成立的可能性是假设 G 成立的可能性的 3 倍. 如果 G 成立时新的证据出现的可能性是 H 成立时的两倍, 那么当新证据出现时, 哪个假设更有可能成立?

9. 你外出度假时, 请邻居给你的病树浇水. 如果没浇水的话, 它死去的概率为 0.8. 如果浇水的话, 它死去的概率为 0.15. 你有 90% 的把握确定邻居记得浇水.

(a) 当你回来时, 树还活着的概率是多大?

(b) 如果树死了, 那么邻居忘记浇水的概率是多大?

10. 一个坛子装有 8 个红球、10 个绿球和 12 个蓝球, 从中随机地取出 6 个球.

(a) 其中至少有一个红球的概率有多大?

(b) 已知其中没有红球的条件下, 在 6 个球中恰有 2 个绿球的概率有多大?

11. 市场上电池有两种牌子. 牌子 C 的电池为合格品的概率为 0.7, 而牌子 D 的电池为合格品的概率为 0.4. 在一盒子中有 8 个牌子为 C 的电池和 6 个牌子为 D 的电池. 现从这盒子中随机地选出一个电池.

(a) 这个电池为合格品的概率有多大?

(b) 已知这个电池为不合格品, 那么它是牌子为 C 的概率有多大?

12. 玛丽亚在一次旅行中带两本书. 假定她喜欢第一本书的概率为 0.6, 喜欢第二本书的概率为 0.5, 两本书都喜欢的概率为 0.4. 求在她不喜欢第一本书的条件下喜欢第二本书的条件概率.

13. 一个坛子装有 20 个红球和 10 个蓝球, 现从坛子中一个一个地取出球.

(a) 求"全部红球被取出"早于"全部蓝球被取出"的概率.

现在假定坛子装有 20 个红球、10 个蓝球和 8 个绿球.

(b) 现在再计算"全部红球被取出"早于"全部蓝球被取出"的概率.

(c) 盒子中的球的颜色消除的顺序为蓝、红、绿的概率有多大?

(d) 计算蓝球是三中颜色球中最早被全部取出的概率.

14. 已知一个硬币抛掷时其正面出现的概率为 0.8. A 在抛掷一次以后, 很匆忙地跑到他的同事 B 处, 告诉 B 他所看到的结果. 然而, 当 A 跑到同事 B 处时, 他以 0.4 的概率忘掉了他抛掷硬币的结果. 于是, 他就随机地挑选一个结果 (正面或反面) 告诉 B. 然而, 他若不忘记的话, 他会将所抛掷的结果告诉 B.

(a) B 被告之硬币正面向上的概率有多大?

(b) B 被告之正确结果的概率有多大?

(c) 已知 B 被告之硬币正面向上的条件下, 硬币的确是正面向上的概率有多大?

15. 某种老鼠, 黑色比棕色占优势 (黑色对应显性基因, 棕色对应隐性基因). 假设某只黑老鼠有个兄弟是棕色的, 但其父母都是黑色的.

(a) 该老鼠是纯黑色的概率是多大? (相对它的基因对是一个黑色, 一个棕色的混合而言.)

(b) 假设当这只老鼠和一只棕色老鼠交配时, 其所有 5 个后代都是黑色的, 此时, 这只老鼠为纯黑老鼠的概率是多大?

16. (a) 在 习题 66b 里, 以开关 1 是否关为条件, 计算电流能从 A 到 B 的概率.

(b) 已知电流能从 A 到 B 的条件下, 开关 3 是合上的条件概率.

17. 在 习题 67 中描述的 $k-n$ 系统, 假定每个元件相互独立且正常工作的概率为 1/2, 求已知系统正常工作的条件下, 元件 1 工作正常的条件概率, 当

(a) $k=1, n=2$. (b) $k=2, n=3$.

18. 琼斯先生为了赢得幸运轮中奖, 设计了一套如下的赌博策略: 当他押注时, 只有当前面 10 次都出现黑色数字时才押注在红色数字上. 他的理由是连续 11 次出现黑色的概率非常小. 你认为他的策略如何?

19. A,B,C 三人同时掷一枚硬币, 掷出正面朝上的概率分别为 P_1, P_2, P_3, 如果有一个人掷出的结果与其他两人不一样, 那么就称他为奇异人. 如果没有出现奇异人, 则继续掷硬币, 直到出现奇异人, 那么 A 被称为奇异人的概率是多大?

20. 假设某次试验有 n 个结果, 结果 i 出现的概率为 $p_i, i=1,\cdots,n$, $\sum_{i=1}^{n} p_i = 1$. 如果观察两次试验, 那么第二次的试验结果大于第一次结果的概率为多大?

21. 如果 A 掷 $n+1$ 枚, B 掷 n 枚均匀硬币, 证明: A 得到的正面朝上数大于 B 得到的正面朝上数的概率为 1/2.

提示: 以每人掷出 n 枚硬币后, 谁具有更多的正面朝上数为条件 (共有三种可能).

22. 对于下列叙述, 试证明或给出反例:

(a) 若 E 独立于 F, 且 E 独立于 G, 则 E 独立于 $F \cup G$;
(b) 若 E 独立于 F, E 独立于 G, 及 $FG = \varnothing$, 则 E 独立于 $F \cup G$;
(c) 若 E 独立于 F, F 独立于 G, 及 E 独立于 FG, 则 G 独立于 EF.

23. 令 A 和 B 为具有正概率的事件, 说明以下叙述是 (i) 必然对; (ii) 必然错; (iii) 可能对.

(a) 若 A 和 B 互不相容, 则它们独立;
(b) 若 A 和 B 独立, 则它们互不相容;
(c) $P(A) = P(B) = 0.6$, 且 A, B 互不相容;
(d) $P(A) = P(B) = 0.6$, 且 A, B 互相独立.

24. 按照发生的概率的大小, 将下列事件排序:

(a) 掷一枚均匀硬币, 正面朝上;
(b) 3 次独立重复试验, 每次成功的概率均为 0.8, 三次都成功;
(c) 7 次独立重复试验, 每次成功的概率均为 0.9, 7 次都成功.

25. 有两个工厂生产收音机, 工厂 A 生产的每台收音机是不合格品的概率为 0.05, 而工厂 B 生产的每台收音机为不合格品的概率为 0.01. 假设你在同一家工厂购买了两台收音机, 且这两台收音机来自 A 厂或 B 厂的概率是相等的. 如果第一台收音机检测后发现是不合格品, 求另一台也是不合格品的条件概率.

26. 证明: 若 $P(A|B) = 1$, 则 $P(B^c|A^c) = 1$.

27. 坛子里开始有一个红球, 一个蓝球. 每步从其中随机地取出一个并同时放入两个同颜色的球. (比如, 若开始取出了红球, 那么在下次取球时, 坛子里有 2 个红球 1 个蓝球.) 利用数学归纳法证明: n 步后, 坛子里正好有 i 个红球的概率为 $1/(n+1), 1 \leqslant i \leqslant n+1$.

28. 共有 $2n$ 张牌, 其中 2 张为“A”. 将这些牌随机分给两位选手, 每人 n 张. 然后两位选手按领牌次序声明自己是否有“A”. 在第一人声称自己有“A”时, 第二人没有“A”的条件概率是多大? 其中

(a) $n = 2$; (b) $n = 10$; (c) $n = 100$.

当 n 趋于无穷大时, 该概率的极限值是多大? 为什么?

29. 市场上一共有 n 种不同的优惠券. 某人收集优惠券, 每次收集一张, 并且各次收集是相互独立的. 已知每次收集到第 i 种优惠券的概率为 $p_i, \sum_{i=1}^{n} p_i = 1$.

(a) 如果已经收集到 n 张优惠券, 那么这 n 张优惠券都不相同的概率是多大?
(b) 现在假设 $p_1 = p_2 = \cdots = p_n = 1/n$. 令 E_i 表示“在收集到的 n 张优惠券里没有第 i 种优惠券”, 对 $P\left(\bigcup_i E_i\right)$ 利用容斥等式证明恒等式

$$n! = \sum_{k=0}^{n} (-1)^k \binom{n}{k} (n-k)^n$$

30. 对任意事件 E 和 F, 证明

$$P(E|E \cup F) \geqslant P(E|F)$$

提示: 以 F 是否发生为条件求 $P(E|E \cup F)$.

第 4 章 随 机 变 量

4.1 随 机 变 量

进行试验时，相对于试验的实际结果而言，通常我们更感兴趣有关试验结果的某些函数. 比如，在掷两枚骰子的游戏中，我们通常更关心两枚骰子的点数之和，而不是各枚骰子的具体值. 也即，我们或许关心骰子点数之和为 7，而不关心实际结果究竟是 (1,6) 或 (2,5)，或 (3,4)，或 (4,3)，或 (5,2)，或 (6,1). 同样，在掷若干枚硬币时，我们或许关心正面朝上的总数，而不关心实际结果有关正面朝上或反面朝上的排列情况. 由上面的例子可以看出，这些感兴趣的量是试验结果的实值函数，我们称之为**随机变量**(random variable). 随机变量是定义在样本空间上试验结果的实值函数.

因为随机变量的取值由试验结果决定，因此我们也将随机变量的可能取值赋予概率.

例 1a 考虑掷 3 枚均匀硬币的试验. 令 Y 表示正面朝上出现的枚数，那么 Y 就是一个随机变量，它的取值为 0,1,2,3 之一，各自概率为

$$
\begin{aligned}
P\{Y=0\} &= P\{(\mathrm{T},\mathrm{T},\mathrm{T})\} = \frac{1}{8} \\
P\{Y=1\} &= P\{(\mathrm{T},\mathrm{T},\mathrm{H}),(\mathrm{T},\mathrm{H},\mathrm{T}),(\mathrm{H},\mathrm{T},\mathrm{T})\} = \frac{3}{8} \\
P\{Y=2\} &= P\{(\mathrm{T},\mathrm{H},\mathrm{H}),(\mathrm{H},\mathrm{T},\mathrm{H}),(\mathrm{H},\mathrm{H},\mathrm{T})\} = \frac{3}{8} \\
P\{Y=3\} &= P\{(\mathrm{H},\mathrm{H},\mathrm{H})\} = \frac{1}{8}
\end{aligned}
$$

此处 H 表示正面朝上，T 表示正面朝下. 因为 Y 的取值必是 0 到 3 之一，故有

$$
1 = P\left(\bigcup_{i=0}^{3}\{Y=i\}\right) = \sum_{i=0}^{3} P\{Y=i\}
$$

关于随机变量取值的概率，其性质与前文介绍的关于事件的概率是一致的. ■

例 1b 一个坛子里装着标有数字 1 到 20 的 20 个球，无放回地从中随机取出 3 个，如果打赌取出的球中至少有一个号码大于等于 17，打赌获胜的概率是多大?

解: 令 X 表示取出的球中最大的号码，那么 X 就是取值为 $3,4,\cdots,20$ 之一的随机变量. 而且，如果我们假定 $\binom{20}{3}$ 种抽取方式中任一种方式都是等可能的，那么

$$P\{X=i\}=\frac{\binom{i-1}{2}}{\binom{20}{3}} \qquad i=3,4,\cdots,20 \tag{1.1}$$

(1.1) 式是这样得到的：事件 $\{X=i\}$ 意味着从 20 个球中抽取一个号码为 i 的球，再从号码为 1 到 $i-1$ 中抽取 2 个球，显然一共有 $\binom{1}{1}\binom{i-1}{2}$ 种抽取方式. 这样就可以得到式 (1.1). 由上式可以得到

$$P\{X=20\}=\frac{\binom{19}{2}}{\binom{20}{3}}=\frac{3}{20}=0.150 \qquad P\{X=19\}=\frac{\binom{18}{2}}{\binom{20}{3}}=\frac{51}{380}\approx 0.134$$

$$P\{X=18\}=\frac{\binom{17}{2}}{\binom{20}{3}}=\frac{34}{285}\approx 0.119 \qquad P\{X=17\}=\frac{\binom{16}{2}}{\binom{20}{3}}=\frac{2}{19}\approx 0.105$$

由于事件 $\{X\geqslant 17\}$ 就是不相交事件 $\{X=i\},i=17,18,19,20$ 的并. 这样，能够赢得打赌的概率就是

$$P\{X\geqslant 17\}\approx 0.105+0.119+0.134+0.150=0.508. \qquad \blacksquare$$

例 1c 设有一枚不均匀的硬币，每次掷硬币时，正面朝上的概率为 p. 现在独立重复地掷硬币，直到出现正面朝上或者已经掷了 n 次为止. 令 X 表示投掷硬币的次数，那么 X 就是一个取值为 $1,2,3\cdots,n$ 之一的随机变量，X 的取值概率为

$$P\{X=1\}=P\{\mathrm{H}\}=p$$

$$P\{X=2\}=P\{(\mathrm{T},\mathrm{H})\}=(1-p)p$$

$$P\{X=3\}=P\{(\mathrm{T},\mathrm{T},\mathrm{H})\}=(1-p)^2p$$

$$\vdots$$

$$P\{X=n-1\}=P\{(\underbrace{\mathrm{T},\mathrm{T},\cdots,\mathrm{T}}_{n-2},\mathrm{H})\}=(1-p)^{n-2}p$$

$$P\{X=n\}=P\{(\underbrace{\mathrm{T},\mathrm{T},\cdots,\mathrm{T}}_{n-1},\mathrm{T}),(\underbrace{\mathrm{T},\mathrm{T},\cdots,\mathrm{T}}_{n-1},\mathrm{H})\}=(1-p)^{n-1}$$

作为验证，注意到

$$P\left(\bigcup_{i=1}^{n}\{X=i\}\right)=\sum_{i=1}^{n}P\{X=i\}=\sum_{i=1}^{n-1}p(1-p)^{i-1}+(1-p)^{n-1}$$

$$= p\left[\frac{1-(1-p)^{n-1}}{1-(1-p)}\right] + (1-p)^{n-1} = 1-(1-p)^{n-1} + (1-p)^{n-1} = 1 \qquad \blacksquare$$

例 1d 某个坛子里有 3 个白球、3 个红球和 5 个黑球. 现从中随机无放回地抽取 3 个球. 假设进行打赌, 抽出每个白球可以赢得 1 元, 每个红球则输掉 1 元. 令 X 表示赢得的总钱数, 那么 X 就是一个取值为 $0, \pm1, \pm2, \pm3$ 之一的随机变量, 各取值概率为

$$P\{X=0\} = \frac{\binom{5}{3} + \binom{3}{1}\binom{3}{1}\binom{5}{1}}{\binom{11}{3}} = \frac{55}{165}$$

$$P\{X=1\} = P\{X=-1\} = \frac{\binom{3}{1}\binom{5}{2} + \binom{3}{2}\binom{3}{1}}{\binom{11}{3}} = \frac{39}{165}$$

$$P\{X=2\} = P\{X=-2\} = \frac{\binom{3}{2}\binom{5}{1}}{\binom{11}{3}} = \frac{15}{165}$$

$$P\{X=3\} = P\{X=-3\} = \frac{\binom{3}{3}}{\binom{11}{3}} = \frac{1}{165}$$

这些概率是怎样算出来的呢? 以 $P\{X=0\}$ 的计算为例, 由于三个球是随机地抽取, 因此这 $\binom{11}{3}$ 种球的组合都是等可能的; 而 $\{X=0\}$ 可以分解为两种情况, 一种是抽出来的球全为黑球, 这种抽取方法共有 $\binom{5}{3}$ 种, 另一种情况是红、白、黑各抽出一个球, 这种抽取方式有 $\binom{3}{1}\binom{3}{1}\binom{5}{1}$ 种, 这样, 我们就得到 $P\{X=0\}$ 的计算公式. 对于 $P\{X=i\}$ 的其他几个公式, 论证是类似的. 下面的公式验证了我们的计算结果:

$$\sum_{i=0}^{3} P\{X=i\} + \sum_{i=1}^{3} P\{X=-i\} = \frac{55+39+15+1+39+15+1}{165} = 1$$

因此, 我们赢钱的概率为

$$\sum_{i=1}^{3} P\{X=i\} = \frac{55}{165} = \frac{1}{3}. \qquad \blacksquare$$

例 1e 设想有 N 种不同的优惠券, 某人收集优惠券, 每次收集一张, 且每种优惠券都以相同的可能性被收集到. 又假定各次收集是相互独立的. 假设某人想收集一

套 N 种优惠券，他的办法是一张一张地收集，直到收集成一套为止 (此时 N 种优惠券每种至少一张). 他所收集到的优惠券的总张数是一个随机变量，记为 T. 与其直接计算 $P(T=n)$, 不如先考虑 T 大于 n 的概率. 为此，先固定 n, 并且定义事件 $A_1, A_2, \cdots, A_N$ 如下：A_j 表示"前 n 张优惠券里没有第 j 种优惠券" $(j=1,\cdots,N)$. 因此

$$\begin{aligned}P\{T>n\}=&P\left(\bigcup_{j=1}^{N}A_j\right)=\sum_j P(A_j)-\sum_{j_1<j_2}\sum P(A_{j_1}A_{j_2})+\cdots\\&+(-1)^{k+1}\sum_{j_1<j_2<\cdots<j_k}\sum\sum P(A_{j_1}A_{j_2}\cdots A_{j_k})\cdots\\&+(-1)^{N+1}P(A_1A_2\cdots A_N)\end{aligned}$$

由于每张优惠券不属于第 j 种的概率为 $(N-1)/N$, 利用各次收集相互独立的假设可得

$$P(A_j)=\left(\frac{N-1}{N}\right)^n$$

而当前n 张优惠券里，既没有第 j_1 种优惠券，也没有第 j_2 种优惠券时，$A_{j_1}A_{j_2}$ 发生，因此 $P(A_{j_1}A_{j_2})=\left(\frac{N-2}{N}\right)^n$. 用类似的推理，可得 $P(A_{j_1}A_{j_2}\cdots A_{j_k})=\left(\frac{N-k}{N}\right)^n$, 这样，对于 $n>0$, 我们有

$$\begin{aligned}P\{T>n\}=&N\left(\frac{N-1}{N}\right)^n-\binom{N}{2}\left(\frac{N-2}{N}\right)^n+\binom{N}{3}\left(\frac{N-3}{N}\right)^n-\cdots\\&+(-1)^N\binom{N}{N-1}\left(\frac{1}{N}\right)^n=\sum_{i=1}^{N-1}\binom{N}{i}\left(\frac{N-i}{N}\right)^n(-1)^{i+1}\end{aligned}\tag{1.2}$$

T 等于 n 的概率可结合下式得到

$$P\{T>n-1\}=P\{T=n\}+P\{T>n\}$$

或等价地，

$$P\{T=n\}=P\{T>n-1\}-P\{T>n\}$$

另外一个值得感兴趣的随机变量是前 n 张优惠券里，优惠券的不同种类数，不妨记为 D_n. 为了计算 $P\{D_n=k\}$, 我们首先把注意力放在一组特定的 k 种优惠券，然后计算我们收集的前 n 张优惠券是由这特定的 k 种优惠券组成的概率. 而这个事件说明所收集到的前 n 张优惠券应当满足

A： 每张都是这k种优惠券之一

B： 这k种优惠券的任一种都在n张优惠券中出现

这样，我们得到

$P\{$收集到的n张优惠券由特定的k种优惠券组成$\}=P(AB)=P(A)P(B|A)$. 由于收集的每张优惠券属于这 k 种之一的概率为 k/N, 因此事件 A 的概率为 $(k/N)^n$. 而且，在给定每张优惠券是所考虑的 k 种之一的条件下，很容易看出在

收集优惠券时, k 种优惠券中的任一种都是等可能地被收集. 因此, 给定 A 发生的条件下 B 恰好是事件"收全一套 k 种优惠券, 所需收集张数小于等于 n", 其概率正好是本例前面讨论的 $\{T \leqslant n\}$ 的概率, 不过优惠券的总数是 k 而不是原来的 N. 利用 (1.2) 式, 将公式中的 N 换成 k, 我们得到 $P(A) = \left(\frac{k}{N}\right)^n, P(B|A) = 1 - \sum_{i=1}^{k-1}\binom{k}{i}\left(\frac{k-i}{k}\right)^n(-1)^{i+1}$. 最后, 由于 k 个种类的集合一共有 $\binom{N}{k}$ 种, 因此

$$P\{D_n = k\} = \binom{N}{k}P(AB) = \binom{N}{k}\left(\frac{k}{N}\right)^n\left[1 - \sum_{i=1}^{k-1}\binom{k}{i}\left(\frac{k-i}{k}\right)^n(-1)^{i+1}\right] \quad ■$$

注释 由于至少要收集到 N 张优惠券才有可能得到一整套, 因此, 如果 $n < N$, 则 $P\{T > n\} = 1$. 因此, 由公式 (1.2) 我们可得到一个非常有意义的组合恒等式: 对于 $1 \leqslant n < N$, 有

$$\sum_{i=1}^{N-1}\binom{N}{i}\left(\frac{N-i}{N}\right)^n(-1)^{i+1} = 1$$

也可写成

$$\sum_{i=0}^{N-1}\binom{N}{i}\left(\frac{N-i}{N}\right)^n(-1)^{i+1} = 0$$

或者, 两边同时乘以 $(-1)^N N^n$, 并令 $j = N - i$, 得

$$\sum_{j=1}^{N}\binom{N}{j}j^n(-1)^{j-1} = 0 \qquad 1 \leqslant n < N \quad ■$$

对于随机变量 X, 如下定义的函数 F

$$F(x) = P\{X \leqslant x\} \qquad -\infty < x < \infty$$

称为 X 的**累积分布函数** (cumulative distribution function), 简称为**分布函数**. 因此, 对任一给定实数 x, 分布函数等于该随机变量小于等于 x 的概率.

假设 $a \leqslant b$, 由于事件 $\{X \leqslant a\}$ 包含于事件 $\{X \leqslant b\}$, 可知前者的概率 $F(a)$ 要小于等于后者的概率 $F(b)$. 换句话说, $F(x)$ 是 x 的非降函数. 第 4.10 节给出了分布函数的更一般的性质.

4.2 离散型随机变量

若一个随机变量最多有可数多个可能取值, 则称这个随机变量为离散型的. 对于一个离散型随机变量 X, 我们定义 X 的**概率分布列**(简称为**分布列**, probability mass function) $p(a)$ 如下:

$$p(a) = P\{X = a\}$$

分布列 $p(a)$ 在最多在可数个 a 上取正值, 也即, 如果 X 的可能值为 $x_1, x_2, \cdots$, 那么

$$p(x_i) \geqslant 0 \qquad i = 1, 2, \cdots$$
$$p(x) = 0 \qquad 所有其他x$$

由于 X 必定取值于 $\{x_1, x_2, \cdots\}$, 这样有

$$\sum_{i=1}^{\infty} p(x_i) = 1$$

用较直观的图形方式, 将 $p(x_i)$ 标在 y 轴上, 将 x_i 标在 x 轴上. 例如, 设 X 的分布列为:

$$p(0) = \frac{1}{4} \qquad p(1) = \frac{1}{2} \qquad p(2) = \frac{1}{4}$$

可表示为图 4.1. 类似地, 在掷两枚均匀骰子的试验中, 令 X 为两枚骰子的点数之和, 则 X 的分布列可用图 4.2 表示.

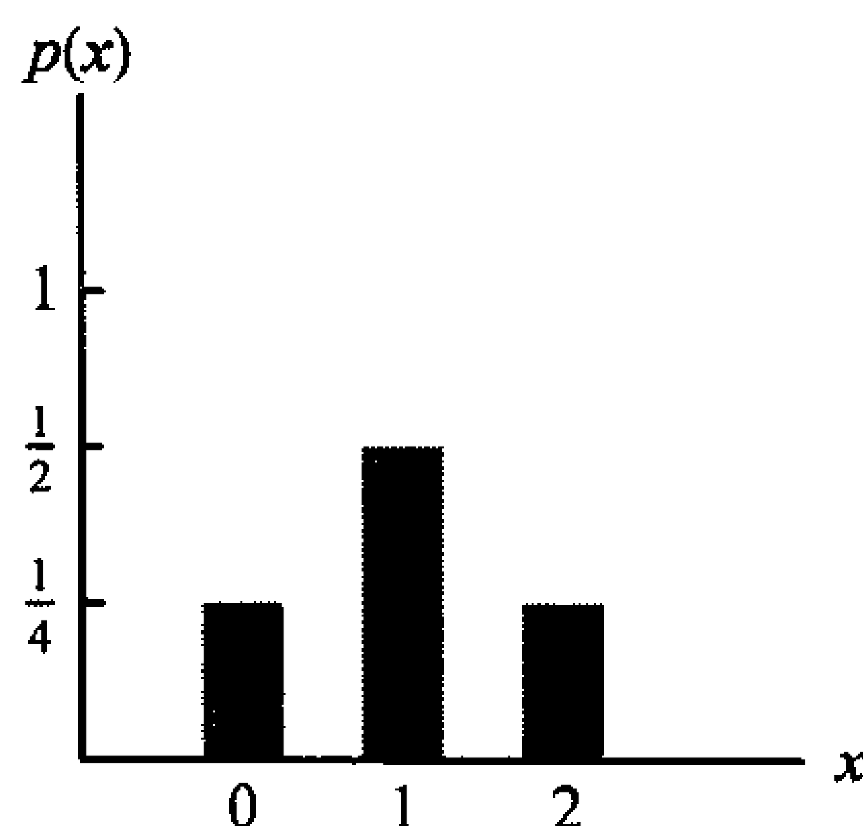


图4.1 分布列1

例 2a 设随机变量 X 的分布列为: $p(i) = c\lambda^i/i\,!, i = 0, 1, 2\cdots$, 其中 λ 为一正数. 求(a) $P\{X=0\}$; (b) $P\{X>2\}$.

解: 因为 $\sum_{i=0}^{\infty} p(i) = 1$, 我们有

$$c\sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = 1$$

又因为 $e^x = \sum_{i=0}^{\infty} x^i/i\,!$, 因此 $ce^{\lambda} = 1$, 也即, $c = e^{-\lambda}$ 这样

(a) $P\{X = 0\} = e^{-\lambda}\lambda^0/0\,! = e^{-\lambda}$

(b) $P\{X > 2\} = 1 - P\{X \leqslant 2\} = 1 - P\{X = 0\} - P\{X = 1\} - P\{X = 2\}$

$$= 1 - e^{-\lambda} - \lambda e^{-\lambda} - \frac{\lambda^2 e^{-\lambda}}{2}$$ ■

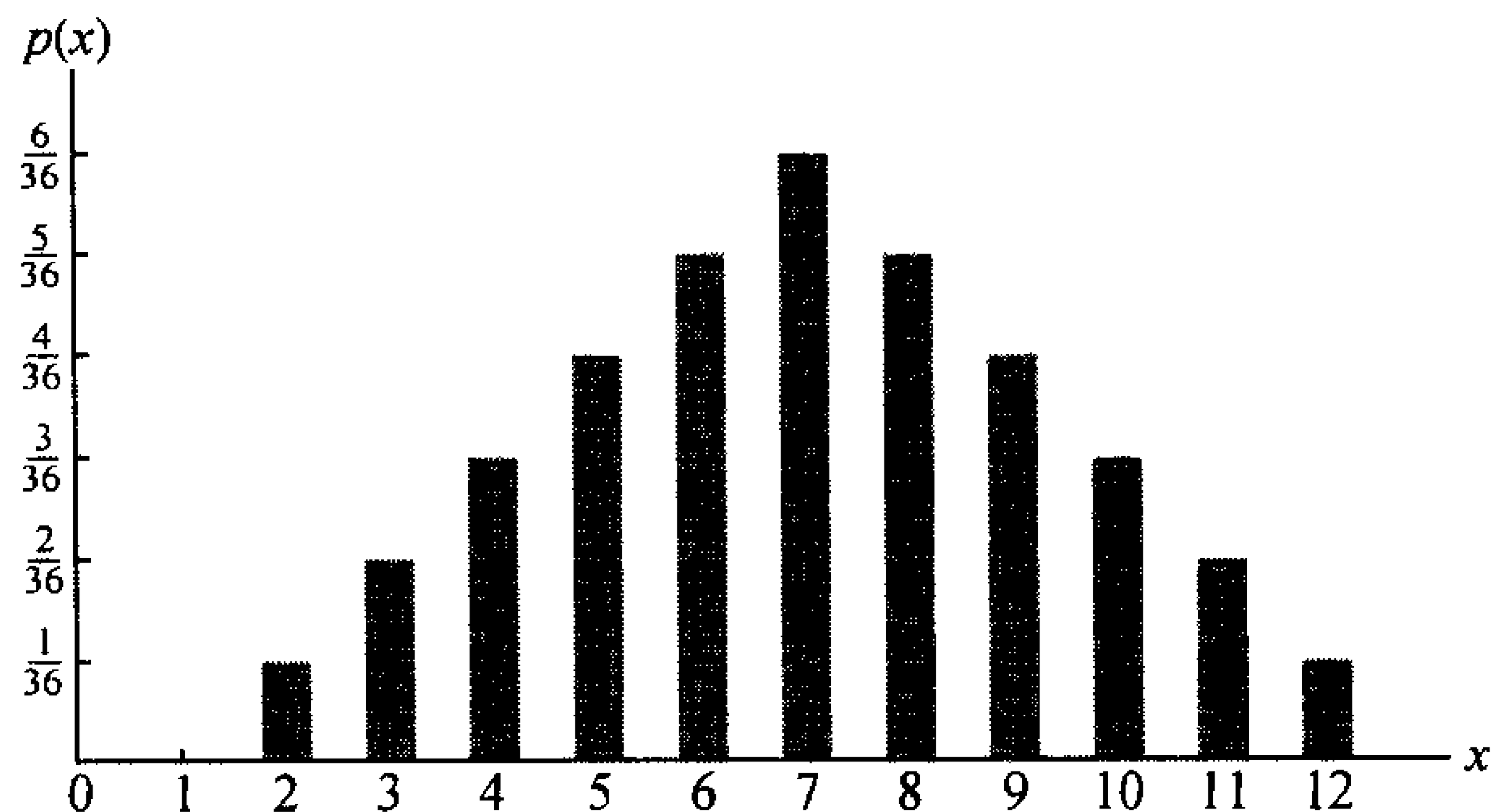


图 4.2 分布列 2

离散型随机变量的分布函数可通过分布列 $p(a)$ 进行计算：

$$F(a)=\sum_{x:x\leqslant a}p(x)$$

如果 X 是一个离散型随机变量，其可能取值为 $x_1,x_2,x_3,\cdots$，其中 $x_1<x_2<x_3<\cdots$，那么它的分布函数是一个阶梯函数．也即，在区间 $[x_{i-1},x_i)$ 上取常数值，且在 x_i 处跳一步，跳跃值为 $p(x_i)$．举例来说，如果 X 的分布列如下：

$$p(1)=\frac{1}{4}\qquad p(2)=\frac{1}{2}\qquad p(3)=\frac{1}{8}\qquad p(4)=\frac{1}{8}$$

那么其分布函数为

$$F(a)=\begin{cases}0 & a<1\\ \frac{1}{4} & 1\leqslant a<2\\ \frac{3}{4} & 2\leqslant a<3\\ \frac{7}{8} & 3\leqslant a<4\\ 1 & 4\leqslant a\end{cases}$$

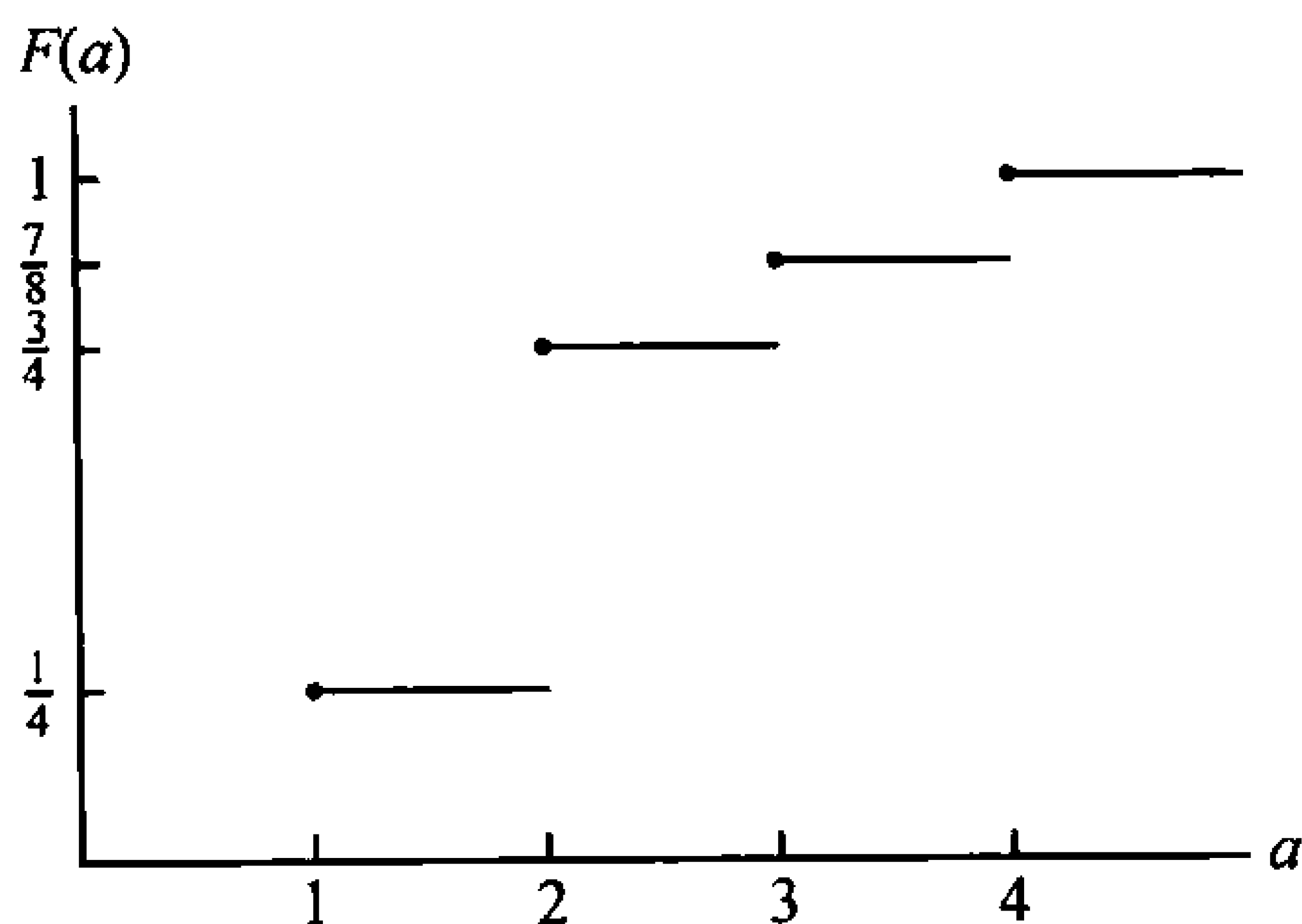


图4.3　分布函数

其图像如图 4.3 所示．

读者应该注意到，$F(a)$ 在 1,2,3,4 处的跳跃值分别等于 X 在该点取值的概率．

4.3　期　　望

概率论里一个非常重要的概念就是随机变量的期望．如果 X 是一个离散型随机变量，并具有分布列 $p(x)$，那么 X 的**期望**(expectation) 或**期望值**(expected value) 记为 $E[X]$，定义如下：

$$E[X]=\sum_{x:p(x)>0}xp(x)$$

也就是说，X 的期望值就是 X 所有可能取值的一个加权平均，每个值的权重就是 X 取该值的概率．例如，如果 X 的分布列如下：

$$p(0)=\frac{1}{2}=p(1)$$

那么

$$E[X]=0\times\frac{1}{2}+1\times\frac{1}{2}=\frac{1}{2}$$

这正是 X 的两个可能取值 0 和 1 在通常意义下的平均值．另外，如果

$$p(0)=\frac{1}{3}\qquad p(1)=\frac{2}{3}$$

那么

$$E[X] = 0 \times \frac{1}{3} + 1 \times \frac{2}{3} = \frac{2}{3}$$

这正是两个可能取值 0 和 1 的加权平均, 此处 1 的权重是 0 的权重的 2 倍, 因为 $p(1) = 2p(0)$.

期望定义的另外一种来源是概率的频率定义. 这个定义 (见第 8 章给出的强大数定律) 指出如果进行无限多次独立重复试验, 那么对任一事件 E, E 发生的次数的比例为 $P(E)$. 现在假设随机变量 X 的可能取值为 $x_1, x_2, \cdots, x_n$, 并且其相应的概率分别为 $p(x_1), p(x_2), \cdots, p(x_n)$, 与此同时, 我们将随机变量 X 解释为一次机会游戏的赢得. 每次游戏, 我们以概率 $p(x_i)$ 赢得 $x_i (i = 1, 2, \cdots, n)$ 个单位. 现在, 利用频率解释, 如果我们连续玩这个游戏, 那么我们赢得 x_i 的频率为 $p(x_i)$, $i = 1, 2, \cdots, n$. 因此我们的平均赢得为

$$\sum_{i=1}^{n} x_i p(x_i) = E[X].$$

例 3a 设 X 表示"掷一枚均匀骰子出现的点数", 求 $E[X]$.

解: 由于 $p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = 1/6$, 这样可得

$$E[X] = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} = \frac{7}{2}$$ ■

例 3b 我们称 I 是事件 A 的示性变量, 如果

$$I = \begin{cases} 1 & \text{如果}A\text{发生} \\ 0 & \text{否则} \end{cases}$$

试求 $E[I]$.

解: 因为 $p(1) = P(A), p(0) = 1 - P(A)$, 这样我们有 $E[I] = P(A)$, 即事件 A 的示性变量的期望就等于事件 A 发生的概率. ■

例 3c 某人参加"答题秀", 一共有问题 1 和问题 2 两个问题. 他可以自行决定回答的顺序. 如果他先回答问题 i, 那么只有回答正确, 他才被允许回答问题 $j (i \neq j)$, 如果回答不正确, 就不允许回答另一问题. 如果他正确回答了问题 i, 他将获得 V_i 元奖励 $(i = 1, 2)$. 这样, 一旦他两道问题都回答正确, 他将获得 $V_1 + V_2$ 元奖励. 如果他能正确回答问题 i 的概率为 $P_i (i = 1, 2)$, 那么他先回答哪个问题才能使得获得奖励的期望值最大化? 设事件 $E_i, i = 1, 2$ 表示"他能正确回答问题 i", 且 E_1, E_2 相互独立.

解: 如果他先回答问题 1, 那么他将获得的奖励如下:

$$\begin{array}{ll} 0 & \text{概率为}1 - P_1 \\ V_1 & \text{概率为}P_1(1 - P_2) \\ V_1 + V_2 & \text{概率为}P_1 P_2 \end{array}$$

因此, 这种情况下他获得的期望奖励是: $V_1P_1(1-P_2)+(V_1+V_2)P_1P_2$. 另一方面, 如果他决定先回答问题 2, 那么获得的期望奖励是: $V_2P_2(1-P_1)+(V_1+V_2)P_1P_2$. 因此, 如果 $V_1P_1(1-P_2)\geqslant V_2P_2(1-P_1)$ 成立, 他应该先回答问题 1, 上式等价于

$$\frac{V_1P_1}{1-P_1}\geqslant\frac{V_2P_2}{1-P_2}$$

这样, 举例来说, 如果他有 60% 的把握答对问题 1, 答对将获得 200 元奖励, 有 80% 的把握答对问题 2, 答对将获得奖励 100 元, 那么他应该选择先回答问题 2, 因为

$$400=\frac{100\times0.8}{0.2}>\frac{200\times0.6}{0.4}=300$$ ■

例 3d 一个学校的 120 名同学分乘 3 辆大客车去听交响乐表演. 第一辆车有 36 名同学, 第二辆有 40 名, 第三辆有 44 名. 到达目的地后, 从 120 名同学中随机抽取一名. 令 X 表示被随机选中的同学所乘坐的车上的同学数, 求 $E[X]$.

解: 随机抽取就意味着 120 名同学被抽中的可能性是一样的, 所以有

$$P\{X=36\}=\frac{36}{120}\qquad P\{X=40\}=\frac{40}{120}\qquad P\{X=44\}=\frac{44}{120}$$

这样,

$$E[X]=36\times\frac{3}{10}+40\times\frac{1}{3}+44\times\frac{11}{30}=\frac{1208}{30}=40.2667$$

另一方面, 一辆客车上的同学数的平均值为 $120/3=40$. 计算表明随机抽取一位同学, 他乘坐的车上的同学数的期望值要大于车上的同学数的平均值. 这是很正常的, 因为一辆车上同学越多, 该车上的同学越容易被抽中, 也即, 同学数多的车所占权重要大于同学数少的车所占权重 (见自检习题 4). ■

注释 期望这一概念可以类比于质量分布的**重心**(Center of gravity) 这一物理概念. 假定有一个离散型随机变量 X 具有分布列 $p(x_i), i\geqslant1$. 我们设想有一根没有重量的竹竿, 在点 x_i 处放有质量 $p(x_i), i\geqslant1$(见图 4.4). 那么能使竹竿保持平衡状态的点就是重心, 对于具有静力学基本知识的读者来说, 很容易找到这个点就是 $E[X]$ ①. ■

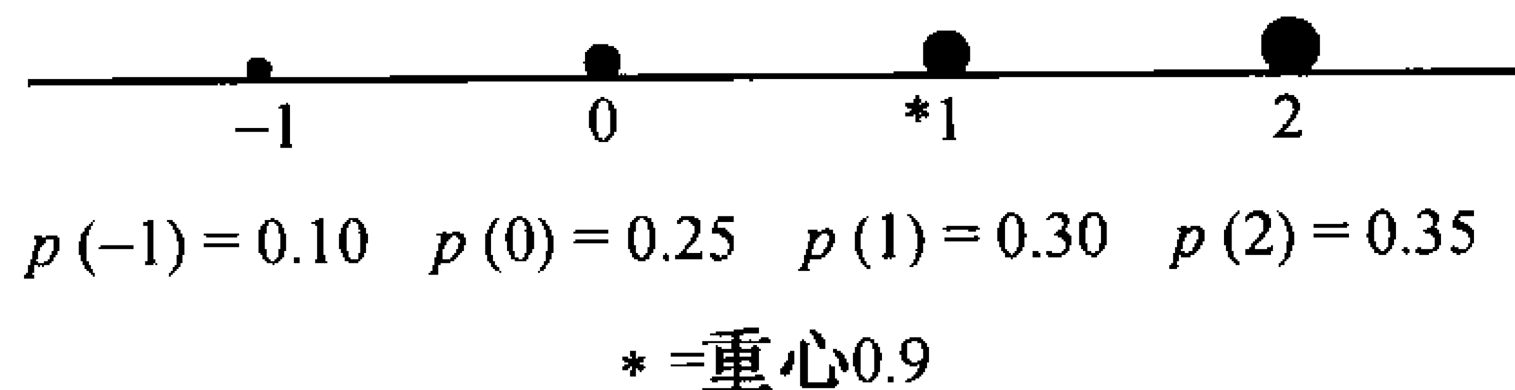


图 4.4 点上的质量

① 为了证明这点, 我们要证明绕 $E[X]$ 的力矩之和等于 0, 也即, 我们要证明 $0=\sum_i(x_i-E[X])P(x_i)$, 而这点很容易得到.

4.4 随机变量函数的期望

如果已知一个随机变量 X 的分布列，如何计算 X 的函数，比如 $g(X)$ 的期望？下面介绍的是一种方法：既然 $g(X)$ 本身也是一个随机变量，它便有自己的分布列，这个分布列可以通过 X 的分布列计算得到. 一旦 $g(X)$ 的分布列得到了，那么就可以根据期望的定义来计算 $g(X)$ 的期望 $E[g(X)]$.

例 4a 令 X 表示取值于 $\{-1,0,1\}$ 的随机变量，其相应的取值概率为：

$$P\{X=-1\}=0.2 \qquad P\{X=0\}=0.5 \qquad P\{X=1\}=0.3$$

计算 $E[X^2]$.

解：令 $Y=X^2$，那么 Y 的分布列如下

$$P\{Y=1\}=P\{X=-1\}+P\{X=1\}=0.5$$
$$P\{Y=0\}=P\{X=0\}=0.5$$

因此，$E[X^2]=E[Y]=1\times 0.5+0\times 0.5=0.5$. 注意到 $0.5=E[X^2]\neq(E[X])^2$ $=0.01$. ■

尽管用上述方法可以根据对随机变量 X 的了解求出一些有关 X 的函数的期望值，但是，对于 $E[g(X)]$ 还有另一种理解方法. 注意到一旦 $X=x$，那么，$g(X)=g(x)$，因此很合理地认为 $E[g(X)]$ 就是 $g(x)$ 的一个加权平均，每个权重就是 $X=x$ 的概率. 这样，以下结论就非常直观.

> **命题 4.1** 如果 X 是一个离散型随机变量，其可能取值为 $x_i, i\geqslant 1$，相应取值概率为 $p(x_i)$，那么对任一实值函数 g，都有
>
> $$E[g(X)]=\sum_i g(x_i)p(x_i)$$

在证明此命题之前，让我们考察例 4a 中的随机变量 X，先看看利用命题所得到的结果是不是与例 4a 的计算结果一致. 利用命题 4.1 可以得到

$$E[X^2]=(-1)^2\times 0.2+0^2\times 0.5+1^2\times 0.3=1\times(0.2+0.3)+0\times 0.5=0.5$$

这个结果与例 4a 的计算结果是一致的.

命题 4.1 的证明：我们的证明方法是将和号 $\sum_i g(x_i)p(x_i)$ 中具有相同 $g(x_i)$ 值的项合并. 特别地，假设 $y_j, j\geqslant 1$，表示 $g(x_i), i\geqslant 1$ 的不同取值，因此可得

$$\begin{aligned}\sum_i g(x_i)p(x_i) &= \sum_j \sum_{i:g(x_i)=y_j} g(x_i)p(x_i) \sum_j \sum_{i:g(x_i)=y_j} y_j P(x_i) \\ &= \sum_j y_j \sum_{i:g(x_i)=y_j} p(x_i)=\sum_j y_j P\{g(X)=y_j\}=E[g(X)]\end{aligned}$$

□

例 4b 某种季节性销售的产品, 如果每卖出一件商品, 可获得纯利润 b 元, 如果季节末仍未卖出, 则每件商品将损失 ℓ 元. 设某百货商店在某个季节的销售量(即卖出商品的件数) 为一随机变量, 其分布列为 $p(i), i \geqslant 0$. 现在商店决定销售旺季前要囤货, 问它要囤多少件才能使得期望利润最大化.

解: 令 X 表示季节销售量, 如果囤货数量为 s, 记利润为 $P(s)$, $P(s)$ 可表示为

$$P(s) = \begin{cases} bX - (s - X)\ell & 如果X \leqslant s \\ sb & 如果X > s \end{cases}$$

因此, 期望利润为

$$\begin{aligned} E[P(s)] &= \sum_{i=0}^{s}[bi - (s-i)\ell]p(i) + \sum_{i=s+1}^{\infty} sbp(i) \\ &= (b+\ell)\sum_{i=0}^{s} ip(i) - s\ell\sum_{i=0}^{s} p(i) + sb\Big[1 - \sum_{i=0}^{s} p(i)\Big] \\ &= (b+\ell)\sum_{i=0}^{s} ip(i) - (b+\ell)s\sum_{i=0}^{s} p(i) + sb = sb + (b+\ell)\sum_{i=0}^{s}(i-s)p(i) \end{aligned}$$

为了得到最佳的 s 值, 我们来看看当 s 增加一个单位时, 利润有什么变化. 利用上述公式得到:

$$\begin{aligned} E[P(s+1)] &= b(s+1) + (b+\ell)\sum_{i=0}^{s+1}(i-s-1)p(i) \\ &= b(s+1) + (b+\ell)\sum_{i=0}^{s}(i-s-1)p(i) \end{aligned}$$

两式相减得

$$E[P(s+1)] - E[P(s)] = b - (b+\ell)\sum_{i=0}^{s} p(i)$$

因此, 如果下列条件满足, 那么囤货数量为 $s+1$ 得到的期望利润会大于囤货数量为 s 的情形:

$$\sum_{i=0}^{s} p(i) < \frac{b}{b+\ell} \tag{4.1}$$

由于公式 (4.1) 的左边随着 s 的增加而增加, 而右边为一常数, 因此不等式对所有的 $s \leqslant s^*$ 总是成立的, 其中 s^* 为满足 (4.1) 式的最大值. 因为

$$E[P(0)] < \cdots < E[P(s^*)] < E[P(s^*+1)] > E[P(s^*+2)] > \cdots$$

这样, 囤货数量达到 s^*+1 时将会使得利润最大化. ■

例 4c 假设你要在两种行动方案之中选择其一, 采取任一种方案都将导致下列 n 个结果之一, 记为 $C_1, \cdots, C_n$. 假设采取了第一种方案, 那么结果 C_i 发生的概率为

$p_i, i = 1, \cdots, n$; 如果采取了第二种方案, 那么结果 C_i 发生的概率为 $q_i, i = 1, \cdots, n$, 其中 $\sum_{i=1}^n p_i = \sum_{i=1}^n q_i = 1$. 下面提供一种选择行动方案的方法. 一开始采取以下方式给各个结果进行赋值: 首先, 确定最坏和最好的结果, 分别称之为 c 和 C, 并给最坏结果 c 赋值 0, 给最好结果 C 赋值 1. 现在考虑其他 $n-2$ 个结果, 不妨设为 C_i. 为了给这些结果赋值, 设想你有以下两种选择: 获得 C_i, 或者采取一随机试验, 使得你获得结果 C 的概率为 u, 获得结果 c 的概率为 $1-u$. 很显然, 你的选择依赖于 u 的取值, 如果 $u=1$, 那么试验取得确定结果 C, 由于 C 为最好的结果, 你肯定认为选择做试验优于选择获得 C_i, 另一方面, 如果 $u=0$, 那么试验结果就是最坏的结果 c, 因此, 这种情形下, 很明显你会认为选择结果 C_i 优于选择做随机试验. 现在, 令 u 从 1 减小到 0, 那么会很合理地认为存在一点, 在这点上, 你认为选择做试验和选择获得 C_i 是一样的. 也即, 在严格的这点上, 这两个选择是没差别的. 那么令这个点所对应的概率 u 为结果 C_i 的值. 换句话说, 你选择 C_i, 或选择做试验, 以概率 u 获得结果 C, 以概率 $1-u$ 获得结果 c, 使得这两者没差别的概率值就是 C_i 的值. 我们称这个概率值为结果 C_i 的*效用*(utility), 记为 $u(C_i)$.

为了确定哪个行动是最优的, 我们要给每个行动估值. 考虑第一个行动方案, 其结果 C_i 发生的概率为 $p_i, i = 1, 2, \cdots, n$, 我们可以认为这个行动方案的结果由一个两步试验决定, 第一步, 随机选择 $1, \cdots, n$, 其相应概率为 p_i, 如果选择了 i, 那么获得结果 C_i, 然而, 由于 C_i 相当于以概率 $u(C_i)$ 获得结果 C, 以概率 $1-u(C_i)$ 获得结果 c. 这样, 两步试验的结果就等价于以下一步试验: 获得结果 C 或者 c, 其中获得 C 的概率为 $\sum_{i=1}^n p_i u(C_i)$. 类似地, 选择第二个行动方案的结果等价于进行试验, 获得结果 C 或者 c, 其中获得 C 的概率为 $\sum_{i=1}^n q_i u(C_i)$. 既然 C 优于 c, 这样第一个试验优于第二个试验, 如果 $\sum_{i=1}^n p_i u(C_i) > \sum_{i=1}^n q_i u(C_i)$, 换句话说, 行动方案可以通过其结果的效用的期望值进行选择. 使得期望效用取得最大值的行动方案是最优的. ■

命题 4.1 的一个简单推论就是以下推论 4.1.

> **推论 4.1** 若 a 和 b 是常数, 则 $E[aX+b] = aE[X] + b$.

证明:

$$E[aX+b] = \sum_{x:p(x)>0}(ax+b)p(x) = a\sum_{x:p(x)>0} xp(x) + b\sum_{x:p(x)>0} p(x) = aE[X] + b \qquad \square$$

随机变量 X 的期望 $E[X]$ 也称为 X 的均值或者一阶矩. $E[X^n], n \geqslant 1$ 称为 X 的 n 阶矩. 由命题 4.1, 可知

$$E[X^n] = \sum_{x:p(x)>0} x^n p(x)$$

4.5 方　　差

给定一随机变量 X 及其分布函数 F, 我们希望能通过定义合适的度量来概括 F 的本质属性, 这也是十分重要和有意义的. 一个比较好的度量是 X 的期望 $E[X]$. 然而, 尽管 $E[X]$ 可以得到 X 取各个可能值的加权平均, 但是它并没有告诉我们任何关于取这些值的对均值的偏离程度, 或者说散布程度等信息. 比如说, 尽管 W, Y, Z 具有分布列如下:

$$W = 0 \quad \text{概率为}1$$

$$Y = \begin{cases} -1 & \text{概率为}1/2 \\ +1 & \text{概率为}1/2 \end{cases} \qquad Z = \begin{cases} -100 & \text{概率为}1/2 \\ +100 & \text{概率为}1/2 \end{cases}$$

这三个随机变量具有相同的期望值 0, 但是 Y 的取值散布程度要远远大于 W (取值为常数) 的取值散布程度. 而且, Z 的取值散布程度也大于 Y 的取值散布程度.

由于我们认为 X 的取值在 $E[X]$ 附近, 这样, 一个合理的度量 X 取值散布程度的方法就是考察 X 的取值与 $E[X]$ 的距离. 一个可能方式就是考虑 $E[|X-\mu|]$, 其中 $\mu = E[X]$. 然而, 在数学上处理这种度量是不方便的, 因此, 更容易处理的度量通常考虑 X 与其均值距离的差的平方的期望, 这样我们就有如下定义.

定义 设 X 的期望为 μ, 则 X 的方差记为 $\mathrm{Var}(X)$, 定义如下:

$$\mathrm{Var}(X) = E[(X-\mu)^2]$$

随机变量 X 的方差 $\mathrm{Var}(x)$ 就是刻画随机变量相对于期望值的散布程度的一个很好的度量. 下面导出 $\mathrm{Var}(X)$ 的另一公式:

$$\begin{aligned}\mathrm{Var}(X) &= E[(X-\mu)^2] = \sum_x (x-\mu)^2 p(x) = \sum_x (x^2 - 2\mu x + \mu^2) p(x) \\ &= \sum_x x^2 p(x) - 2\mu \sum_x x p(x) + \mu^2 \sum_x p(x) \\ &= E[X^2] - 2\mu^2 + \mu^2 = E[X^2] - \mu^2\end{aligned}$$

也即

$$\boxed{\mathrm{Var}(X) = E[X^2] - (E[X])^2}$$

X 的方差等于 X^2 的期望值减去 X 的期望值的平方. 这也是实践中最方便计算 $\mathrm{Var}(X)$ 的方法.

例 5a 掷一枚均匀骰子, 记 X 表示掷出的点数, 计算 $\mathrm{Var}(X)$.

解: 在例 3a 中已经得到 $E[X]=7/2$, 利用命题 4.1, 得

$$E[X]^2 = 1^2\times\frac{1}{6}+2^2\times\frac{1}{6}+3^2\times\frac{1}{6}+4^2\times\frac{1}{6}+5^2\times\frac{1}{6}+6^2\times\frac{1}{6}=\frac{1}{6}\times 91$$

因此,

$$\mathrm{Var}(X)=\frac{91}{6}-\left(\frac{7}{2}\right)^2=\frac{35}{12}$$ ■

对于任意常数 a 和 b, 下面的等式是十分有用的:

$$\mathrm{Var}(aX+b)=a^2\mathrm{Var}(X)$$

为了证明上式, 令 $\mu=E[X]$, 注意到推论 4.1 的结果 $E[aX+b]=a\mu+b$, 因此有

$$\mathrm{Var}(aX+b)=E[(aX+b-a\mu-b)^2]=E[a^2(X-\mu)^2]=a^2E[(X-\mu)^2]=a^2\mathrm{Var}(X)$$

注释 (a) 在力学中, 类比于均值是质量分布的重心, 方差代表了惯性矩.

(b) $\mathrm{Var}(X)$ 的平方根称为 X 的*标准差*(Standard deviation), 记为 $\mathrm{SD}(X)$, 也即 $\mathrm{SD}(X)=\sqrt{\mathrm{Var}(X)}$.

离散型随机变量通常根据其分布列进行分类, 下面的几节将要介绍几种常见的类型.

4.6 伯努利随机变量和二项随机变量

考虑一个试验, 其结果分为两类, 或者成功, 或者失败. 如果我们令

$$X=\begin{cases}1 & \text{当试验结果为成功时}\\ 0 & \text{当试验结果为失败时}\end{cases}$$

那么 X 的分布列如下:

$$p(0)=P\{X=0\}=1-p \qquad p(1)=P\{X=1\}=p \tag{6.1}$$

其中 p, $0\leqslant p\leqslant 1$ 就是每次试验成功的概率.

一个随机变量 X 称为伯努利随机变量(起源于瑞士数学家詹姆士 · 伯努利), 如果其分布列由 (6.1) 式给出, 其中 $p\in(0,1)$.

现在设进行 n 次独立重复试验, 每次试验成功的概率为 p, 失败的概率为 $1-p$. 若以 X 表示 n 次试验中成功的次数, 那么 X 称为参数为 (n,p) 的*二项随机变量*(binomial). 因此, 伯努利随机变量也称为参数为 $(1,p)$ 的二项随机变量.

参数为 (n,p) 二项随机变量的分布列为

$$p(i)=\binom{n}{i}p^i(1-p)^{n-i} \qquad i=0,1,\cdots,n \tag{6.2}$$

(6.2) 式的成立原因如下：首先注意到某特定包含 i 个成功和 $n-i$ 个失败的 n 个结果的序列的概率为 $p^i(1-p)^{n-i}$(由试验的独立性). 又由于 n 个结果的序列里, 包含 i 个成功和 $n-i$ 个失败的序列一共有 $\binom{n}{i}$ 个, 举例来说, 如果 $n=4, i=2$, 那么包含两次成功两次失败的试验结果序列一共有 $\binom{4}{2}=6$ 种, 也即：(s,s,f,f),(s,f,s,f),(s,f,f,s), (f,s,s,f),(f,s,f,s) 及 (f,f,s,s), 其中 (s,s,f,f) 表示前两次成功, 后两次失败. 既然每个结果序列的概率均为 $p^2(1-p)^2$, 那么在 4 次独立重复试验中, 2 次成功的概率为 $\binom{4}{2}p^2(1-p)^2$.

由二项式定理知

$$\sum_{i=0}^{\infty} p(i) = \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} = [p+(1-p)]^n = 1$$

例 6a 掷 5 枚均匀的硬币, 假定掷各枚硬币所得的结果是相互独立的, 求掷出的 5 枚硬币中正面朝上枚数的分布列.

解： 令 X 表示掷 5 枚硬币正面朝上 (成功) 的硬币枚数, 则 X 就是一个参数为 $(n=5, p=1/2)$ 的二项随机变量, 因此, 由公式 (6.2), 可得：

$$P\{X=0\} = \binom{5}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^5 = \frac{1}{32} \qquad P\{X=1\} = \binom{5}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^4 = \frac{5}{32}$$

$$P\{X=2\} = \binom{5}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^3 = \frac{10}{32} \qquad P\{X=3\} = \binom{5}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^2 = \frac{10}{32}$$

$$P\{X=4\} = \binom{5}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^1 = \frac{5}{32} \qquad P\{X=5\} = \binom{5}{5}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^0 = \frac{1}{32}$$

■

例 6b 某工厂生产螺钉, 已知每一颗螺钉为残次品的概率为 0.01, 并且各螺钉是否残次品是相互独立的. 工厂以 10 颗为一盒出售螺钉, 并且承诺每盒里最多只有一个残次品, 否则该盒作退货处理. 问售出的产品中退货的比例.

解： 令 X 表示一盒中有缺陷的螺钉的数量, 那么 X 就是一个服从二项分布的随机变量, 参数为 $(10, 0.01)$, 因此, 任一箱将要退回的概率为

$$1 - P\{X=0\} - P\{X=1\} = 1 - \binom{10}{0}(0.01)^0(0.99)^{10} - \binom{10}{1}(0.01)^1(0.99)^9 \approx 0.004$$

因此, 将要有 0.4% 的盒子被退回. ■

例 6c 以下介绍的赌博方法称为“运气轮”, 在世界各地的狂欢节或赌场十分流行. 赌徒押注于 1 到 6 之间某一个数, 然后庄家掷 3 枚骰子, 如果赌徒押的数出现 $i, i=1,2,3$ 次, 那么他将赢得 i 单位. 反之, 如果赌徒押的数没出现, 他将损失 1 单位. 问这个赌博对赌徒是否公平? (实际上, 这个赌博经常是转一个轮子, 当轮子

停下来时, 指针会指向某结果, 其结果会显示 1 到 6 之间的三个数字, 但是从数学上来说, 这两者是等价的.)

解: 我们假设骰子都是均匀的, 而且掷出的点数相互独立, 那么赌徒押的数出现的次数就是一个二项随机变量, 其参数为 $(3, 1/6)$, 因此, 令 X 表示赌徒赢得的数目, 我们有

$$P\{X=-1\}=\binom{3}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^3=\frac{125}{216}\quad P\{X=1\}=\binom{3}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^2=\frac{75}{216}$$

$$P\{X=2\}=\binom{3}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^1=\frac{15}{216}\quad P\{X=3\}=\binom{3}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^0=\frac{1}{216}$$

为了计算这赌博方式对赌徒是否公平, 我们先计算 $E[X]$, 利用以上概率可以得到

$$E[X]=\frac{-125+75+30+3}{216}=\frac{-17}{216}$$

因此, 如果长期赌下去, 每 216 局, 赌徒将要输掉 17 个单位. ■

接下来的例子我们考虑由孟德尔 (G. Mendel, 1822—1884) 发展的遗传理论的最简单的一种形式.

例 6d 考虑人类的某一特别属性 (比如眼睛的颜色, 或是否左撇子等) 是由他的一对基因决定的, 以 d 表示显性基因, 而 r 表示隐性基因, 则有 dd 基因的是纯显性的, 有 rr 的是纯隐性的, 而有 dr 的就是混合性. 纯显性的和混合性都显露出显性基因决定的特征. 孩子从其父母身上各遗传得到 1 个基因, 如果一对混合型父母总共有 4 个孩子, 问其中 3 个孩子具有显性基因所决定的特征的概率有多大?

解: 我们假定每个孩子独立地从他的父母方各遗传得到 1 个基因, 因此, 当父母均为混合型时, 孩子具有基因 dd,rr,dr 的概率分别是 $1/4, 1/4, 1/2$. 而具有基因 dd,dr 的孩子具有显性基因所决定的特征, 因此, 孩子具有显性基因所决定的特征的概率为 $3/4$, 这样, 他们的 4 个孩子中, 具有显性特征的孩子数的分布是二项分布, 参数为 $n=4, p=3/4$. 那么所求概率为:

$$\binom{4}{3}\left(\frac{3}{4}\right)^3\left(\frac{1}{4}\right)^1=\frac{27}{64}$$ ■

例 6e 某陪审团的审判由 12 名陪审员参加. 为宣判被告有罪, 必须其中至少有 8 名陪审员投票认为他有罪. 假设陪审员的判断是相互独立的, 且每位陪审员作出正确判断的概率为 θ. 问这个陪审团作出正确判决的概率有多大?

解: 上述问题是无法解的, 因为还缺少足够的信息. 例如, 若被告是无罪的, 则陪审团作出正确判决的概率为

$$\sum_{i=5}^{12}\binom{12}{i}\theta^i(1-\theta)^{12-i}$$

而若被告有罪, 正确判决的概率为

$$\sum_{i=8}^{12}\binom{12}{i}\theta^i(1-\theta)^{12-i}$$

因此, 如果已知被告的确有罪的概率为 α, 那么以他是不是有罪为条件, 我们得到陪审团作出正确判决的概率为

$$\alpha\sum_{i=8}^{12}\binom{12}{i}\theta^i(1-\theta)^{12-i}+(1-\alpha)\sum_{i=5}^{12}\binom{12}{i}\theta^i(1-\theta)^{12-i}$$ ■

例 6f 一个通讯系统由 n 个元件组成, 各个元件是否工作正常是相互独立的, 并且各个元件正常工作的概率均为 p. 若在系统中, 至少有一半的元件工作正常, 那么整个系统就有效.

(a) p 取何值时, 5 个元件的系统比 3 个元件的系统更有可能有效?

(b) 一般来说, 什么时候 $2k+1$ 个元件的系统比 $2k-1$ 个元件的系统更有效?

解: (a) 正常工作的元件数是一个服从参数为 (n,p) 的二项分布的随机变量, 那么 5 个元件的系统有效的概率为

$$\binom{5}{3}p^3(1-p)^2+\binom{5}{4}p^4(1-p)+p^5$$

而三个元件的系统有效的概率为

$$\binom{3}{2}p^2(1-p)+p^3$$

因此, 以下条件成立时, 5 个元件的系统比 3 个元件的系统更有效:

$$10p^3(1-p)^2+5p^4(1-p)+p^5>3p^2(1-p)+p^3$$

化简为 $3(p-1)^2(2p-1)>0$, 即 $p>1/2$.

(b) 一般来说, 当 (且仅当) $p>1/2$ 时, $2k+1$ 个元件的系统比 $2k-1$ 个元件的系统有效. 为了证明这点, 考虑 $2k+1$ 元件的系统, 令 X 表示"前 $2k-1$ 个元件中工作正常的元件的数目", 那么

$$P_{2k+1}(\text{系统有效})=P\{X\geqslant k+1\}+P\{X=k\}(1-(1-p)^2)+P\{X=k-1\}p^2$$

上式之所以成立是基于事件"$2k+1$ 元件的系统有效"可以写成下列三个互不相容的事件的并:

(i) $X\geqslant k+1$;

(ii) $X=k$ 而且剩下的 2 个元件中至少有 1 个工作正常;

(iii) $X=k-1$ 而且剩下的 2 个元件都工作正常.

由于

$$P_{2k-1}(\text{ 工作有效}) = P\{X \geqslant k\} = P\{X = k\} + P\{X \geqslant k+1\}$$

可得

$$\begin{aligned}
&P_{2k+1}(\text{ 工作有效}) - P_{2k-1}(\text{ 工作有效}) \\
&\quad = P\{X = k-1\}p^2 - (1-p)^2 P\{X = k\} \\
&\quad = \binom{2k-1}{k-1} p^{k-1}(1-p)^k p^2 - (1-p)^2 \binom{2k-1}{k} p^k (1-p)^{k-1} \\
&\quad = \binom{2k-1}{k} p^k (1-p)^k [p - (1-p)] \qquad \text{因为} \binom{2k-1}{k-1} = \binom{2k-1}{k} \\
&\quad > 0 \Leftrightarrow p > 1/2
\end{aligned}$$

■

4.6.1 二项随机变量的性质

现在我们来考察参数为 (n, p) 的二项随机变量的性质, 先来计算其期望和方差.

$$E[X^k] = \sum_{i=0}^{n} i^k \binom{n}{i} p^i (1-p)^{n-i} = \sum_{i=1}^{n} i^k \binom{n}{i} p^i (1-p)^{n-i}$$

利用恒等式 $i\binom{n}{i} = n\binom{n-1}{i-1}$, 可以得到

$$\begin{aligned}
E[X^k] &= np \sum_{i=1}^{n} i^{k-1} \binom{n-1}{i-1} p^{i-1} (1-p)^{n-i} \\
&= np \sum_{j=0}^{n-1} (j+1)^{k-1} \binom{n-1}{j} p^j (1-p)^{n-1-j} \qquad \text{令} j = i-1 \\
&= npE[(Y+1)^{k-1}]
\end{aligned}$$

其中, Y 是一个参数为 $(n-1, p)$ 的二项随机变量. 在上面的公式中, 令 $k=1$, 就可以得到 $E[X] = np$, 也即, 如果试验每次成功的概率为 p , 那么 n 次独立重复试验的成功总次数的期望值等于 np. 在前面的公式中令 $k=2$, 再利用上面得到的关于二项随机变量的期望公式, 可得 $E[X^2] = npE[Y+1] = np[(n-1)p+1]$, 因为 $E[X] = np$, 所以可得

$$\text{Var}(X) = E[X^2] - (E[X])^2 = np[(n-1)p+1] - (np)^2 = np(1-p)$$

总结以上论述, 可得如下结论.

> 如果 X 是一个参数为 (n, p) 的二项随机变量, 那么
>
> $$E[X] = np \qquad \text{Var}(X) = np(1-p)$$

接下来的命题表明, 二项分布的分布列一开始是递增, 后来逐渐递减.

命题 6.1 如果 X 是一个参数为 (n,p) 的二项随机变量, 其中 $0<p<1$. 那么当 k 从 0 到 n 时, $P\{X=k\}$ 一开始单调递增, 然后一直单调递减, 它在 $k=[(n+1)p]$ 时取最大值 (记号 $[X]$ 表示小于等于 X 的最大的整数).

证明: 为证明这命题, 我们考虑 $P\{X=k\}/P\{X=k-1\}$, 对于给定的 k 值, 判定以下比值与 1 的大小关系.

$$\frac{P\{X=k\}}{P\{X=k-1\}}=\frac{\dfrac{n!}{(n-k)!\,k!}p^k(1-p)^{n-k}}{\dfrac{n!}{(n-k+1)!\,(k-1)!}p^{k-1}(1-p)^{n-k+1}}=\frac{(n-k+1)p}{k(1-p)}$$

因此, $P\{X=k\}\geqslant P\{X=k-1\}$ 当且仅当 $(n-k+1)p\geqslant k(1-p)$, 或者, 等价于 $k\leqslant(n+1)p$. 这样, 命题得到了证明. □

图 4.5 为参数为 $(10,1/2)$ 的二项随机变量的概率分布图.

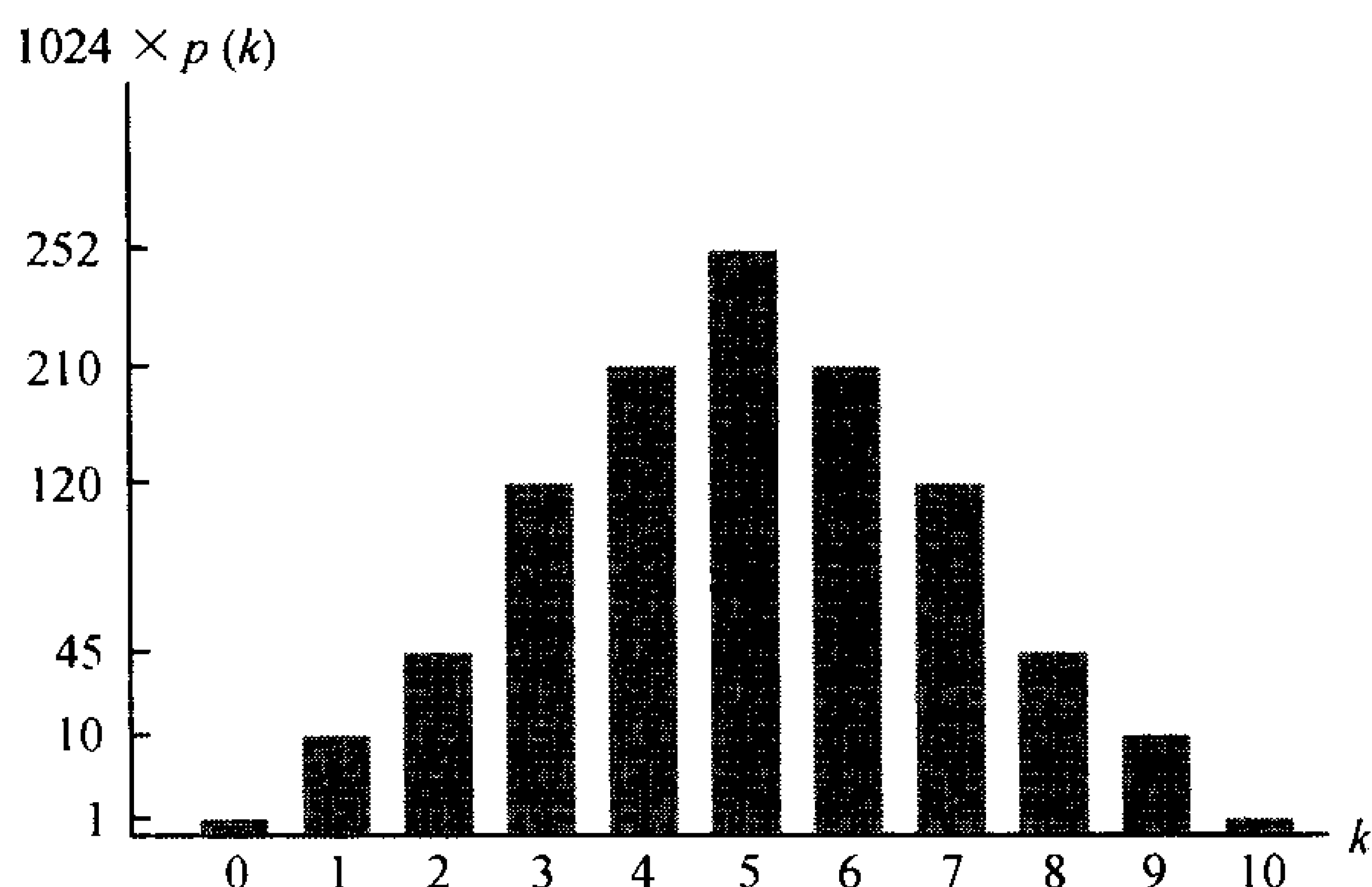


图 4.5 $p(k)=\binom{10}{k}\left(\frac{1}{2}\right)^{10}$

例 6g 在美国总统选举中, 若候选人在一个州里获得了多数的选票, 那么该候选人就赢得了分配给该州的全部选举团票. 选举团票数正比于该州的人口数. 设该州的人口数为 n, 选举团票数约为 nc(实际上, 这个数约为 $nc+2$, 该州的每一个众议员有一张选票, 众议员人数正比于该州的人口数. 而该州的参议员也有一张选票, 而一个州的参议员只有 2 名). 现在我们要计算一个选民的平均权力. 所谓平均权力是与你的一票起的关键作用有关的量. 例如, 在某次会议上, 你有决定权, 你的权力就大. 那么, 在选举中, 你的权力体现在哪里? 设想这一州一共有 $n=2k+1$ 人 (n 为偶数时, 情况是类似的), 候选人共 2 人, 除你之外其余的 $2k$ 人对这两个候选人的态度一样, 即选票中有 k 个人支持候选人甲, 另外 k 个人支持乙. 在这种情况下, 你的一票是关键的. 若你支持甲, 甲得胜, 支持乙, 乙得胜. 现在, 我们假定州里的 $2k$ 个人的选择是相互独立的, 并且每个选民选择甲或乙的概率都是 1/2. 这个

时候一个选民的选票起关键作用的概率为:

$$P\{公民的选票起关键作用\} = \binom{2k}{k}\left(\frac{1}{2}\right)^k\left(\frac{1}{2}\right)^k = \frac{(2k!)}{k!k!2^{2k}}$$

现在我们利用斯特林公式,

$$k! \sim k^{k+1/2}\mathrm{e}^{-k}\sqrt{2\pi}$$

此处, $a_k \sim b_k$ 表示当 $k \to \infty$ 时 $a_k/b_k \to 1$. 这样我们得到

$$P\{公民的选票起关键作用\} \sim \frac{(2k)^{2k+1/2}\mathrm{e}^{-2k}\sqrt{2\pi}}{k^{2k+1}\mathrm{e}^{-2k}(2\pi)2^{2k}} = \frac{1}{\sqrt{k\pi}}$$

当你的选票起关键作用时, 你将影响到 nc 张选举团票, 这样, 在一个人口为 n 的州里的选民, 平均起来影响到多少张选举团票呢? 我们用平均权力这个指标.

$$\begin{aligned} 平均权力 &= nc \cdot P\{你的一票是关键的\} + 0 \cdot P\{你的一票不起关键作用\} \\ &\sim \frac{nc}{\sqrt{n\pi/2}} = c\sqrt{2n/\pi} \end{aligned}$$

这样看来, 你的平均权力与州的人口数的平方根成正比, 在大州中一票的平均权力比小州中的一票的平均权力大. ■

4.6.2 计算二项分布函数

设 X 是一个参数为 (n,p) 的二项随机变量, 计算分布函数

$$P\{X \leqslant i\} = \sum_{k=0}^{i}\binom{n}{k}p^k(1-p)^{n-k} \qquad i = 0, 1, \cdots, n$$

的方法是利用在证明命题 6.1 得到的如下 $P\{X = k+1\}$ 和 $P\{X = k\}$ 之间的关系:

$$P\{X = k+1\} = \frac{p}{1-p}\frac{n-k}{k+1}P\{X = k\} \tag{6.3}$$

例 6h 令 X 为一参数为 $n = 6, p = 0.4$ 的二项随机变量. 从 $P\{X = 0\} = 0.6^6 = 0.0467$ 开始, 利用递推公式 (6.3) 可得

$$P\{X = 0\} = 0.6^6 \approx 0.0467 \qquad P\{X = 1\} = \frac{4}{6} \times \frac{6}{1}P\{X = 0\} \approx 0.1866$$

$$P\{X = 2\} = \frac{4}{6} \times \frac{5}{2}P\{X = 1\} \approx 0.3110 \qquad P\{X = 3\} = \frac{4}{6} \times \frac{4}{3}P\{X = 2\} \approx 0.2765$$

$$P\{X = 4\} = \frac{4}{6} \times \frac{3}{4}P\{X = 3\} \approx 0.1382 \qquad P\{X = 5\} = \frac{4}{6} \times \frac{2}{5}P\{X = 4\} \approx 0.0369$$

$$P\{X = 6\} = \frac{4}{6} \times \frac{1}{6}P\{X = 5\} \approx 0.0041$$

■

很早就有人写了一个利用递推公式 (6.3) 计算二项随机变量的分布函数的程序. 为了计算 $P\{X \leqslant i\}$ 必须先计算 $P\{X = 0\}$[①], 然后再利用递推公式计算 $P\{X = 1\}$

① 原书之义是先计算 $P\{X = i\}$, 再计算 $P\{X = i-1\}$ 等, 实际应先计算 $P\{X = 0\}$, 再计算 $P\{X = 1\}$ 等.——译者注

和 $P\{X=2\}$, 等等. 这样的程序可在网站上找到. 运行它时, 直接输入二项分布的参数 n 和 p 以及一个值 i, 程序就会计算出参数为 (n,p) 的二项随机变量小于或等于 i 的概率.

历 史 注 记

瑞士数学家雅克 · 伯努利 (Jacques Bernoulli, 1654—1705) 首次研究独立重复试验 (每次成功率为 p). 在他去世后的第 8 年 (1713 年), 他侄子尼克拉斯出版了伯努利的著作《推测术》. 在书中, 伯努利指出了如果这样的试验次数足够大, 那么成功次数所占的比例以概率 1 接近 p .

雅克 · 伯努利是这个最著名的数学家庭的第一代. 在后来的三代里, 一共有 8 到 12 个伯努利, 在概率论、统计学和数学上作出了杰出的基础性贡献. 知道其具体数目比较困难, 一方面是有好几人的名字相同 (比如, 雅克的兄弟让有两个儿子分别叫雅克和让), 另一方面是有几个伯努利在不同的地方有不同的名字. 比如, 我们刚说的雅克 (有时也写成 Jaques) 有时叫雅可布 (也写成 Jacob) 或詹姆士 · 伯努利. 但不管如何, 他们的成果和影响都是非凡的. 正如巴赫家族之于音乐, 伯努利家族在数学界是非常有名的家族!

例 6i　设 X 是一个服从参数为 $n=100, p=0.75$ 的二项随机变量, 求 $P\{X=70\}$ 和 $P\{X\leqslant 70\}$.

解: 如图 4.6 所示.

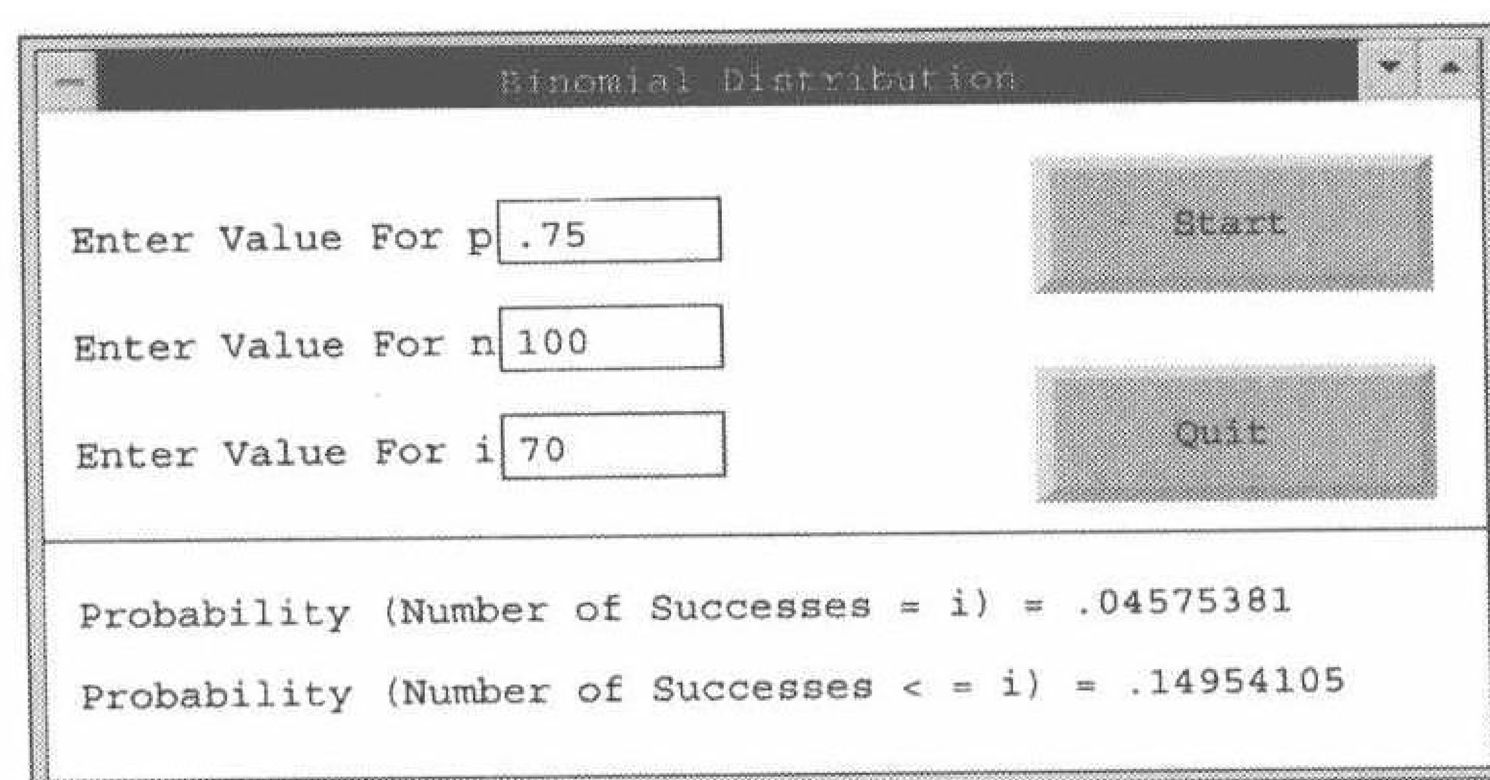


图 4.6　统计软件计算概率

4.7　泊松随机变量

一个取值为 $0,1,2\cdots$ 之一的随机变量称为服从参数为 λ 的泊松随机变量(Poisson), 如果对某一 $\lambda>0$, 有

$$p(i)=P\{X=i\}=\mathrm{e}^{-\lambda}\frac{\lambda^i}{i!}\qquad i=0,1,2,\cdots \tag{7.1}$$

公式 (7.1) 定义了一个分布列, 因为

$$\sum_{i=1}^{\infty} p(i) = \mathrm{e}^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i\,!} = \mathrm{e}^{-\lambda}\mathrm{e}^{\lambda} = 1$$

泊松分布是 S.D. 泊松在他所著的关于概率论在诉讼、刑事审讯等方面应用的书中提出的, 这本书于 1837 年出版, 法文书名叫*Recherchés sur la probabilité des jugements en matière criminelle et en matière civile.*

泊松分布在各种各样的领域中有着非常广泛的应用, 这是由于当 n 足够大, p 充分小, 而使得 np 保持适当的大小时, 以 (n,p) 为参数的二项随机变量可近似看作泊松分布. 为证明这点, 设 X 是一个服从参数为 (n,p) 的二项随机变量, 并且记 $\lambda = np$, 这样

$$\begin{aligned} P\{X=i\} &= \frac{n\,!}{(n-i)\,!\,i\,!} p^i (1-p)^{n-i} = \frac{n\,!}{(n-i)\,!\,i\,!} \left(\frac{\lambda}{n}\right)^i \left(1-\frac{\lambda}{n}\right)^{n-i} \\ &= \frac{n(n-1)\cdots(n-i+1)}{n^i} \frac{\lambda^i}{i\,!} \frac{(1-\lambda/n)^n}{(1-\lambda/n)^i} \end{aligned}$$

由于对充分大的 n 和适当的 λ, 有

$$\left(1-\frac{\lambda}{n}\right)^n \approx \mathrm{e}^{-\lambda} \qquad \frac{n(n-1)\cdots(n-i+1)}{n^i} \approx 1 \qquad \left(1-\frac{\lambda}{n}\right)^i \approx 1$$

因此, 有

$$P\{X=i\} \approx \mathrm{e}^{-\lambda} \frac{\lambda^i}{i!}$$

也就是说, 独立重复进行 n 次试验, 每次成功率为 p, 当 n 充分大, 而 p 足够小, 使得 np 保持适当的话, 那么成功的次数近似服从参数为 $\lambda = np$ 的泊松分布, 这个 λ 值 (以后将要证明这就是成功次数的期望值) 通常凭经验确定.

以下例子中的随机变量通常都服从泊松分布 (也即满足公式 (7.1)):

1. 一本书里一页或若干页中的印刷错误;
2. 某地区居民活到 100 岁的人数;
3. 一天中拨错电话号码的总数;
4. 一家便利店里每天卖出狗粮饼干的盒数;
5. 某一天进入某邮局的顾客数;
6. 一年中联邦司法系统中空缺位置数;
7. 某放射性材料在一定时期内放射出来的 α-粒子数.

还有其他大量的随机变量, 都因为相同的原因近似服从泊松分布, 也即, 因为泊松分布与二项分布很近似. 例如, 我们认为某一页上任一字母出现印刷错误的概率 p 是一个很小的数, 因此, 这一页上总的印刷错误近似服从参数为 $\lambda = np$ 的泊松分布, 其中 n 是该页上的字母数. 类似地, 我们还认为某个地区某人能活到 100 岁的概率很小; 同样, 进入某家商店的顾客购买一袋狗粮的概率也可以认为是很小的数, 等等.

例 7a 假设本书的某一页上的印刷错误数服从参数为 1/2 的泊松分布, 计算该页上至少有一处错误的概率.

解: 令 X 表示该页上的错误数, 我们有

$$P\{X \geqslant 1\} = 1 - P\{X = 0\} = 1 - \mathrm{e}^{-1/2} \approx 0.393 \quad \blacksquare$$

例 7b 假设某台机器生产出来的零件为残次品的概率为 0.1, 计算有 10 个这样的零件的样本中至多有一个残次品的概率有多大?

解: 所求概率为 $\binom{10}{0} \times 0.1^0 \times 0.9^{10} + \binom{10}{1} \times 0.1^1 \times 0.9^9 = 0.7361$, 而利用泊松分布近似可得该概率值为 $\mathrm{e}^{-1}+\mathrm{e}^{-1} \approx 0.7358$. ■

例 7c 考虑这样一个试验: 记录 1 克放射性物质在 1 秒内放出的 α 粒子数. 如果从过去的经验得知, 这个数目的平均值为 3.2, 问放出的 α 粒子数不超过 2 的概率的较好的近似值是多少?

解: 设想这 1 克放射性物质由 n 个原子组成 (n 相当大), 每个原子在所考虑的 1 秒内蜕变并放出一个 α 粒子的概率为 $3.2/n$, 于是我们可看到, 放射出的 α 粒子数近似服从参数为 $\lambda = 3.2$ 的泊松分布. 因此, 所求的概率为

$$P\{X \leqslant 2\} = \mathrm{e}^{-3.2} + 3.2\mathrm{e}^{-3.2} + \frac{3.2^2}{2}\mathrm{e}^{-3.2} \approx 0.3799 \quad \blacksquare$$

在计算参数为 λ 的泊松随机变量的期望和方差之前, 回顾它近似于参数为 n 和 p(其中 n 很大, p 很小, $\lambda = np$) 的二项随机变量, 而这个二项随机变量的期望值为 $np = \lambda$, 方差为 $np(1-p) = \lambda(1-p) \approx \lambda$(因为 p 很小). 这样, 看起来好像泊松随机变量的期望和方差都等于其参数 λ. 下面我们来证明这一点.

$$\begin{aligned} E[X] &= \sum_{i=0}^{\infty} \frac{i\mathrm{e}^{-\lambda}\lambda^i}{i\,!} = \lambda \sum_{i=1}^{\infty} \frac{\mathrm{e}^{-\lambda}\lambda^{i-1}}{(i-1)\,!} = \lambda\mathrm{e}^{-\lambda}\sum_{j=0}^{\infty}\frac{\lambda^j}{j\,!} \qquad \text{令} j = i-1 \\ &= \lambda \qquad \text{因为} \sum_{j=0}^{\infty}\frac{\lambda^j}{j\,!} = \mathrm{e}^{\lambda} \end{aligned}$$

因此, 泊松随机变量 X 的期望等于其参数 λ, 为了计算其方差, 先计算 $E[X^2]$.

$$\begin{aligned} E[X^2] &= \sum_{i=0}^{\infty} \frac{i^2\mathrm{e}^{-\lambda}\lambda^i}{i\,!} = \lambda \sum_{i=1}^{\infty} \frac{i\mathrm{e}^{-\lambda}\lambda^{i-1}}{(i-1)\,!} = \lambda \sum_{j=0}^{\infty} \frac{(j+1)\mathrm{e}^{-\lambda}\lambda^j}{j\,!} \qquad \text{令} j = i-1 \\ &= \lambda\left[\sum_{j=0}^{\infty}\frac{j\mathrm{e}^{-\lambda}\lambda^j}{j\,!} + \sum_{j=0}^{\infty}\frac{\mathrm{e}^{-\lambda}\lambda^j}{j\,!}\right] = \lambda(\lambda+1) \end{aligned}$$

其中, 最后一个等式成立是因为第一项就是参数为 λ 的泊松随机变量的期望, 而第二项就是该随机变量取各个值的概率之和. 因此, 由于我们已经得到 $E[X] = \lambda$, 可得 $\mathrm{Var}(X) = E[X^2] - (E[X])^2 = \lambda$. 即

泊松随机变量的期望和方差都等于其参数 λ.

我们指出了参数为 np 的泊松分布是对 n 次独立重复试验 (每次成功的概率为 p) 中成功次数的较好的近似, 其中给定条件 n 很大而 p 很小. 事实上, 在试验并不独立, 但是弱相依条件下仍是比较好的近似. 例如, 匹配问题 (第 2 章例 5m), 其中 n 个人随机地从他们的帽子中取一顶帽子, 考虑恰好拿着自己的帽子的人数. 可认为这 n 个选择就是 n 次试验, 其中第 i 次成功就是第 i 个人拿到了自己的帽子, $i = 1, \cdots, n$. 定义事件 $E_i, i = 1, \cdots, n$ 如下:

$$E_i = \{\text{第}i\text{次试验成功}\}$$

很容易看出 $P\{E_i\} = 1/n$, 且$P\{E_i|E_j\} = 1/(n-1), j \neq i$, 那么很合理地认为成功的次数近似服从参数为 $n \times 1/n = 1$ 的泊松分布, 事实上, 这一点已经在第 2 章例 5m 得到了证明.

现在给出第二种关于试验为弱相依情形下泊松近似的阐述, 考虑第 2 章例 5i. 在该例中, 假设有 n 个人, 每个人在一年 365 天内任一天过生日的概率都相同, 现在的问题是计算 n 个人生日各不相同的概率. 我们曾用组合学知识计算了该概率, 并计算出当 $n = 23$ 时该概率小于 1/2.

我们可以利用泊松近似来给出上述概率的近似值. 设想我们进行一系列试验, 对于不同的 i 和 j(两个人), 称试验 i, j 为成功, 如果 i 和 j 生日相同. 如果我们令 E_{ij} 表示事件“试验 i, j 成功”, 那么这 $\binom{n}{2}$ 个事件 $E_{ij}, 1 \leqslant i < j \leqslant n$ 并不独立 (见理论 习题 21), 但其相依性却很弱 (事实上, 对于不同的 i, j, k, l, E_{ij} 与 E_{kl} 是相互独立的, 见理论 习题 21). 由于 $P(E_{ij}) = 1/365$, 因此, 假设成功次数近似服从参数为 $\binom{n}{2}/365 = n(n-1)/730$ 的泊松分布是很合理的. 因此

$$P\{\text{没有 2 个人生日相同}\} = P\{0\text{个成功}\} \approx \exp\left\{\frac{-n(n-1)}{730}\right\}$$

现在计算该概率小于 1/2 的 n 的值, 注意到

$$\exp\left\{\frac{-n(n-1)}{730}\right\} \leqslant \frac{1}{2}$$

等价于

$$\exp\left\{\frac{n(n-1)}{730}\right\} \geqslant 2$$

两边取对数, 可得 $n(n-1) \geqslant 730 \ln 2 \approx 505.997$. 这样, 可以得到解: $n = 23$. 这与第 2 章例 5i 的结论是一致的.

现在考虑我们想要得到没有 3 个人同一天生日的概率, 这时用组合学知识去解决就变得相当困难了, 但仍可以用简单的方法得到一个较好的近似. 首先, 设想

我们对 $\binom{n}{3}$ 个 $i,j,k,1\leqslant i<j<k\leqslant n$ 做一次试验, 试验称为成功的, 如果 i,j,k 这 3 人生日相同. 如上所述, 我们知道成功数近似为泊松随机变量, 参数为

$$\binom{n}{3}P\{i,j,k\text{ 这 3 人生日相同}\}=\binom{n}{3}\left(\frac{1}{365}\right)^2=\frac{n(n-1)(n-2)}{6\times 365^2}$$

因此,

$$P\{\text{没有 3 人生日相同}\}\approx\exp\left\{\frac{-n(n-1)(n-2)}{799\ 350}\right\}$$

该概率值小于 1/2, 若 n 满足 $n(n-1)(n-2)\geqslant 799\ 350\ln 2\approx 554\ 067.1$, 等价于 $n\geqslant 84$, 因此, 当人数超过 84 时, 至少有 3 人生日相同的概率超过 1/2.

因此要使事件发生的数量近似服从泊松分布, 并没有必要要求各个事件发生的概率相同, 只要这些概率都较小即可. 下面就是泊松范例.

泊松范例 考虑 n 个事件, 每个事件发生的概率为 $p_i,i=1,\cdots,n$. 如果所有 p_i 都很小, 且试验或者独立, 或者至多弱相依, 那么事件发生次数近似服从参数为 $\sum_{i=1}^{n}p_i$ 的泊松分布.

接下的例子不但应用了泊松范例, 而且阐述了前面介绍的一系列技巧.

例 7d (最大游程的长度) 抛掷硬币 n 次, 假定各次抛掷是相互独立的, 每次抛掷正面朝上的概率为 p. 出现连续 k 次正面朝上的概率有多大?

解: 首先应用泊松范例逼近这个概率. 对于 $i=1,\cdots,n-k+1$, 令 H_i 表示"第 $i,i+1,\cdots,i+k-1$ 次抛掷硬币均为正面朝上". 此时, 连续 k 次正面朝上的概率就是至少有一个 H_i 发生的概率. 由于 H_i 是"第 $i,i+1,\cdots,i+k-1$ 次抛掷硬币均为正面朝上", $P(H_i)=p^k$. 当 p^k 很小时, H_i 发生的次数应该近似具有泊松分布. 但是, 这是不对的, 因为尽管各事件 H_i 发生的概率很小, 而某些事件之间的依赖性很大, 影响了泊松逼近的精度. 在第 $1,\cdots,k$ 次抛掷硬币的结果都是正面朝上的条件下, 第 $2,\cdots,k+1$ 次抛掷硬币的结果都是正面朝上的概率等于第 $k+1$ 次抛掷的结果为正面朝上的概率. 即 $P(H_2|H_1)=p$, 这个概率比 $P(H_2)$ 大得多.

为了应用泊松逼近, 对于 $i=1,\cdots,n-k$, 令 E_i 为"第 $i,i+1,\cdots,i+k-1$ 次抛掷的结果都是正面朝上, 而第 $i+k$ 次抛掷为反面朝上". 而 E_{n-k+1} 为"第 $n-k+1,\cdots,n$ 次抛掷都是正面朝上", 这样我们得到

$$P(E_i)=p^k(1-p)\qquad i\leqslant n-k$$
$$P(E_{n-k+1})=p^k$$

这样, 当 p^k 很小时, $P(E_i)$ 都是小概率事件, 对于 E_i 和 E_j, 当它们所涉及的试验没有相重的时候, $P(E_i|E_j)=P(E_i)$, 如他们涉及的试验有相重部分, $P(E_i|E_j)=0$. 在这两种情形下, 条件概率都可以为是无条件的. 令 N 表示 E_i 发生的次数, N 的分布应该近似为泊松分布, 其期望为

$$E[N]=\sum_{i=1}^{n-k+1}P(E_i)=(n-k)p^k(1-p)+p^k$$

不存在 k 次连续正面朝上的充要条件为 $N=0$. 因此

$$P\{\text{不存在}k\text{次连续的正面朝上}\}=P\{N=0\}\approx\exp\{-(n-k)p^k(1-p)-p^k\}$$

现在令 L_n 为“n 次试验中连续出现正面的最大次数”, 亦即 L_n 为 n 次试验中的出现正面的最大游程的长度. 易知 $L_n<k$ 的充要条件是试验序列中没有 k 次连续正面朝上的一段. 因此, 利用上式

$$P\{L_n<k\}\approx\exp\{-(n-k)p^k(1-p)-p^k\}$$

现在假定硬币是均匀的, 即 $p=1/2$, 此时上式变成

$$P\{L_n<k\}\approx\exp\left\{-\frac{n-k+2}{2^{k+1}}\right\}\approx\exp\left\{-\frac{n}{2^{k+1}}\right\}$$

上面最后的近似式利用了 $\exp\{(k-2)/2^{k+1}\}\approx 1$ 或 $(k-2)/2^{k+1}\approx 0$. 令 $j=\ln_2 n$, 并假定 j 为整数, 令 $k=j+i$, 有

$$\frac{n}{2^{k+1}}=\frac{n}{2^j2^{i+1}}=\frac{1}{2^{i+1}}$$

显然,

$$P\{L_n<j+i\}\approx\exp\{-(1/2)^{i+1}\}$$

由此可知

$$\begin{aligned}P\{L_n=j+i\}&=P\{L_n<j+i+1\}-P\{L_n<j+i\}\\&\approx\exp\{-(1/2)^{i+2}\}-\exp\{-(1/2)^{i+1}\}\end{aligned}$$

例如,

$P\{L_n<j-3\}\approx e^{-4}\approx 0.0183$

$P\{L_n=j-3\}\approx e^{-2}-e^{-4}\approx 0.1170$

$P\{L_n=j-2\}\approx e^{-1}-e^{-2}\approx 0.2325$

$P\{L_n=j-1\}\approx e^{-1/2}-e^{-1}\approx 0.2387$

$P\{L_n=j\}\approx e^{-1/4}-e^{-1/2}\approx 0.1723$

$P\{L_n=j+1\}\approx e^{-1/8}-e^{-1/4}\approx 0.1037$

$P\{L_n=j+2\}\approx e^{-1/16}-e^{-1/8}\approx 0.0569$

$P\{L_n=j+3\}\approx e^{-1/32}-e^{-1/16}\approx 0.0298$

$P\{L_n\geqslant j+4\}\approx 1-e^{-1/32}\approx 0.0308$

由上式看出, 不管 n 有多大, n 次试验的最大的正面朝上的游程的长度在 $\ln_2(n)-3$ 和 $\ln_2(n)+1$ 之间的概率大约为 0.86.

现在我们要导出在 n 次抛掷硬币试验中, 连续 k 次出现正面的概率 (每次抛掷硬币正面朝上的概率为 p). 利用记号 L_n 及 E_i, 我们有

$$P(L_n \geqslant k) = P\{\text{在}n\text{次试验中出现连续}k\text{个正面向上}\} = P\Big(\bigcup_{i=1}^{n-k+1} E_i\Big)$$

利用事件和的概率的容斥恒等式,

$$P\Big(\bigcup_{i=1}^{n-k+1} E_i\Big) = \sum_{r=1}^{n-k+1} (-1)^{r+1} \sum_{i_1<\cdots<i_r} P(E_{i_1}\cdots E_{i_r})$$

令 S_i 表示与事件 E_i 相关联的试验号的集合. 例如, $S_1 = \{1,\cdots,k+1\}$. 现在考虑 $E_1,\cdots,E_{n-k}$ 中 r 个事件的交的概率 (把事件 E_{n-k+1} 排除在考虑之列), 即考虑 $P(E_{i_1}\cdots E_{i_r}) i_1 < \cdots < i_r < n-k+1$, 若 $S_{i_1},\cdots,S_{i_r}$ 中任何两个集合有相交的情况, $P(E_{i_1}\cdots E_{i_r}) = 0$. 如果两两不相交, 则 $E_{i_1},\cdots,E_{i_r}$ 是相互独立的, 这样

$$P(E_{i_1}\cdots E_{i_r}) = \begin{cases} 0 & \text{若}S_{i_1},\cdots,S_{i_r}\text{中有两个相交的情况} \\ p^{rk}(1-p)^r & \text{若}S_{i_1},\cdots,S_{i_r}\text{互不相交} \end{cases}$$

现在我们要确定 $i_1 < \cdots < i_r < n-k+1$ 中使得 $S_{i_1},\cdots,S_{i_r}$ 互不相交的组数. 首先注意到, 对于每个集合 S_{i_j}, $j = 1,\cdots,r$, 对应了 $k+1$ 次抛掷硬币. 而这些集合又不相交, 一共对应于 $r(k+1)$ 次抛掷硬币. 现在考虑 r 个相同的字母 a 和 $n-r(k+1)$ 个相同的字母 b 的排列, 其中每个 a 对应于一个集合 S_{r_j}, 每个 b 对应于 $S_{r_j}, j = 1,\cdots,r$ 以外的一个指标. 现在假定这些 a 和 b 已经排好一顺序. 每一个排列对应于不相交的指标集 $S_{i_1},\cdots,S_{i_r}$ 的一种选择. 排列的第一个 a 前面的 b 的个数代表 S_{i_1} 之前的试验的次数. 若第一个 a 前面有 i_1-1 个 b, 那么 S_{i_1} 刚好由 $\{i_1, i_1+1,\cdots,i_1+k\}$ 组成, 而排在第一个 a 与第二个 a 之间的 b 的个数, 刚好对应于 S_{i_1} 之后, S_{i_2} 之前的试验次数, $\cdots$. 由于这些 a, b 的排列共有 $\binom{n-rk}{r}$ 个. 每一个排列对应于一个不相交的 $S_{i_1},\cdots,S_{i_r}$ 的一种选择, 这样

$$\sum_{i_1<\cdots<i_r<n-k+1} P(E_{i_1}\cdots E_{i_r}) = \binom{n-rk}{r} p^{rk}(1-p)^r$$

注意到在事件和的容斥等式里边的相应的求和公式为

$$\sum_{i_1<\cdots<i_r} P(E_{i_1}\cdots E_{i_r})$$

我们将这个和号进行分解[①]

$$\sum_{i_1<\cdots<i_r} P(E_{i_1}\cdots E_{i_r}) = \sum_{i_1<\cdots<i_r<n-k+1} P(E_{i_1}\cdots E_{i_r}) + \sum_{i_1<\cdots<i_{r-1}} P(E_{i_1}\cdots E_{i_{r-1}} E_{n-k+1})$$

① 原文并没有这个和号分解, 译者为便于读者理解, 添加了此内容.——译者注

$$= \binom{n-rk}{r} p^{rk}(1-p)^r + \sum_{i_1<\cdots<i_{r-1}} P(E_{i_1}\cdots E_{i_{r-1}}E_{n-k+1})$$

在上式右边第二个和号内的各项的计算和第一个和号的计算是一样的. 当 $S_{i_1},\cdots,S_{i_{r-1}},S_{n-k}$ 中某两个集合相交时, $P(E_{i_1}\cdots E_{i_{r-1}}E_{n-k+1})=0$. 当 $S_{i_1},\cdots,S_{i_{r_1}},S_{n-k+1}$ 两两不相交时,

$$P(E_{i_1}\cdots E_{i_{r-1}}E_{n-k+1}) = [p^k(1-p)]^{r-1}p^k = p^{kr}(1-p)^{r-1}$$

这样,

$$\sum_{i_1<\cdots<i_{r-1}} P(E_{i_1}\cdots E_{i_{r-1}}E_{n-k+1}) = K\cdot p^{n-r}(1-p)^{r-1}$$

其中 K 是不相交子集类 $S_{i_1},\cdots,S_{i_{r-1}},S_{n-k}$ 的数目. 这个数目等于 $(r-1)$ 个 a 和 $n-(r-1)(k+1)-k$ 个 b 的合在一起的排列数. 这个排列数为 $\binom{n-rk}{r-1}$. 将所得到的公式代入 $P(L_n\geqslant k)$ 的表达式, 得到

$$P(L_n\geqslant k) = \sum_{r=1}^{n-k+1}(-1)^{r+1}\left[\binom{n-rk}{r}+\frac{1}{p}\binom{n-rk}{r-1}\right]p^{kr}(1-p)^r$$

在上式中我们规定 $\binom{m}{j}=0, m<j$.

从计算的角度看存在更有效的方法计算上述概率, 其方法是导出一个递推公式. 令 A_n 表示"在 n 次抛掷硬币的试验中出现连续 k 次正面朝上"的事件, 记 $P_n=P(A_n)$. 记 F_j 表示"n 次试验中第一次反面朝上在第 j 次抛掷硬币时出现", 用 H 表示"前 k 次都是正面朝上", $F_1,F_2,\cdots,F_k,H$ 形成一个互不相容的完备组. 我们有

$$P(A_n) = \sum_{j=1}^{k} P(A_n|F_j)P(F_j) + P(A_n|H)P(H)$$

由于第一次反面朝上是在第 j 次试验出现, $j<k$. 而我们要出现 k 次连续正面, 因此前 j 次试验不起作用, 相当于试验重新开始, 于是 $P(A_n|F_j)=P_{n-j}$, 又由于 $P(A_n|H)=1$, 我们得到

$$P_n = P(A_n) = \sum_{j=1}^{k} P_{n-j}P(F_j) + P(H) = \sum_{j=1}^{k} P_{n-j}p^{j-1}(1-p) + p^k$$

由于 $P_j=0, j<k$ 及 $P_k=p^k$, 我们利用上式递推算出 $P_{k+1},P_{k+2},\cdots$, 直至算出 P_n. 例如, 我们希望计算出在 4 次抛掷一枚均匀硬币中出现 2 个连续正面朝上的概率, $k=2$, $P_1=0, P_2=(1/2)^2=1/4$, 利用公式

$$P_n = \sum_{j=1}^{k} P_{n-j}\left(\frac{1}{2}\right)^j + \left(\frac{1}{2}\right)^k$$

得到

$$P_3 = P_2\left(\frac{1}{2}\right) + P_1\left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 = \frac{3}{8} \quad P_4 = P_3\left(\frac{1}{2}\right) + P_2\left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 = \frac{1}{2}$$

这显然是正确的. 因为 4 次抛掷硬币, 连续两次出现正面的情况为: hhhh, hhht, hhth, hthh, thhh, hhtt, thht 和 tthh, 共有 8 种情况, 每种情况出现的概率为 1/16. ■

泊松分布的另一应用表现在这样的情形中, "事件" 发生在某些时间点上. 这种事件的例子有: 发生一次地震, 某人进入特定地点 (如银行、邮局、加油站等), 爆发一次战争等. 我们假设这样的事件发生在一列 (随机) 时间点上, 并设存在某个正的常数 λ 使得如下条件成立.

1. 在任意长度为 h 的时间区间内, 正好发生一个事件的概率彼此相同, 都等于 $\lambda h + o(h)$, 其中 $o(h)$ 表示任何满足 $\lim_{h\to 0} f(h)/h = 0$ 的函数 $f(h)$. [例如 $f(h) = h^2$ 是 $o(h)$, 而 $f(h) = h$ 不是 $o(h)$.]

2. 在任意长度为 h 的时间区内发生 2 个或更多个事件的概率非常小, 等于 $o(h)$.

3. 对于任意确定的自然数 n 与非负整数 $j_1, j_2, \cdots, j_n$, 以及任意 n 个互不相交的时间区间, 若以 E_i 表示 "在第 i 个时间区内上述事件正好发生 j_i 次", 则 $E_1, E_2 \cdots, E_n$ 相互独立.

粗略地说, 条件 1 与条件 2 说明, 当 h 比较小时, 在长度为 h 的区间内正好发生 1 个事件的概率等于 λh 加上某个比 h 更小的量, 而事件发生多于一次的概率就是一个比 h 更小的量. 条件 3 说明, 在一个时间区间内无论发生了什么, 对另一个与它不相交的区间 (从概率意义上) 没有影响.

在条件 1, 2 与 3 成立的假设下, 我们证明, 在任意长度为 t 的时间区间内, 事件发生的次数是以 λt 为参数的泊松随机变量. 为此, 我们考虑区间 $[0, t]$, 并以 $N(t)$ 表示 "这个区间内事件发生的次数". 为求出 $P\{N(t) = k\}$ 的表达式, 先将区间 $[0, t]$ 等分为 n 个互不相交且长度为 t/n 的子区间 (图 4.7).

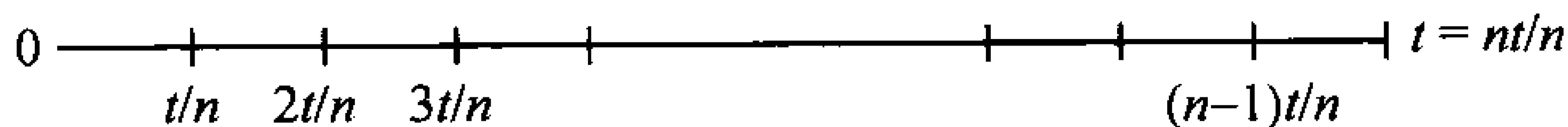


图 4.7 对区间 [0,t] 等分

我们有

$$\begin{aligned} & P\{N(t) = k\} \\ =& P\{n\text{个子区间中某}k\text{个正好各含 1 个事件而其余}n-k\text{个子区间各含 0 个事件}\} \\ & + P\{N(t) = k\text{且至少 1 个子区间含多于 1 个事件}\} \end{aligned} \tag{7.2}$$

显然, 等式 (7.2) 是因为事件 $\{N(t)=k\}$ 等于右边两个互不相容事件之并. 以 A 与 B 分别表示 (7.2) 式右边这两个互不相容事件, 便得到

$$\begin{aligned}
P(B) &\leqslant P\{\text{至少一个子区间包含多于 1 个事件}\} \\
&= P\{\bigcup_{i=1}^{n} \text{第}i\text{个子区间包含多于 1 个事件}\} \\
&\leqslant \sum_{i=1}^{n} P\{\text{第}i\text{个子区间包含多于 1 个事件}\} \qquad \text{用布尔不等式} \\
&= \sum_{i=1}^{n} o\left(\frac{t}{n}\right) = n\,o\left(\frac{t}{n}\right) = t\left[\frac{o(t/n)}{t/n}\right]
\end{aligned}$$

因对任何 t, 当 $n\to\infty$ 时, $t/n\to 0$, 从而由 $o(h)$ 的定义可得, 当 $n\to\infty$ 时, $o(t/n)/(t/n)\to 0$. 因此, 当 $n\to\infty$ 时,

$$P(B)\to 0 \tag{7.3}$$

另一方面, 由于条件 1 与条件 2 蕴涵①

$$\begin{aligned}
&P\{\text{在长度为}h\text{的区间内有 0 个事件发生}\} \\
&= 1-[\lambda h+o(h)+o(h)] = 1-\lambda h-o(h)
\end{aligned}$$

又由独立性条件 3 可得

$$\begin{aligned}
&P\{A\} \\
&=P\{n\text{个子区间中某}k\text{个正好各含 1 个事件而其余}n-k\text{个子区间各含 0 个事件}\} \\
&=\binom{n}{k}\left[\frac{\lambda t}{n}+o\left(\frac{t}{n}\right)\right]^{k}\left[1-\left(\frac{\lambda t}{n}\right)-o\left(\frac{t}{n}\right)\right]^{n-k}
\end{aligned}$$

但因当 $n\to\infty$ 时,

$$n\left[\frac{\lambda t}{n}+o\left(\frac{t}{n}\right)\right] = \lambda t + t\left[\frac{o(t/n)}{t/n}\right]\to\lambda t$$

故采用与证明二项随机变量的泊松近似相同的方法可证, 当 $n\to\infty$ 时,

$$P(A)\to \mathrm{e}^{-\lambda t}\frac{(\lambda t)^k}{k\,!} \tag{7.4}$$

因此, 由 (7.2),(7.3),(7.4) 式, 令 $n\to\infty$, 我们得到

$$P\{N(t)=k\} = \mathrm{e}^{-\lambda t}\frac{(\lambda t)^k}{k\,!} \qquad k=0,1,2,\cdots \tag{7.5}$$

---

① 两个形如 $o(h)$ 的函数之和仍为 $o(h)$, 这是因为, 若 $\lim\limits_{h\to 0} f(h)/h = \lim\limits_{h\to 0} g(h)/h = 0$, 则 $\lim\limits_{h\to 0}[f(h)+g(h)]/h = 0$.

这样, 如果事件的发生满足条件 1, 2, 3, 则在任何固定的长度为 t 的时间区间内, 事件发生的次数是以 λt 为参数的泊松随机变量. 这时, 我们称事件是按速率为 λ 的泊松过程发生的. 数 λ 可解释为单位时间内事件发生的速率. 它一定是由经验确定的常数.

上述讨论阐明了为什么泊松随机变量通常可作为诸如下列各种现象的很好的近似.

1. 发生在某固定时间间隔内地震的次数;

2. 每年爆发战争的次数;

3. 在某固定周期内从一个热阴极放射出的电子数;

4. 某人寿保险公司的保险客户在某一时间区间内死亡的个数.

例 7e 美国西部发生地震的次数符合上述假设 1, 2, 3, 且以 1 周为单位时间, 速率 $\lambda = 2$(即地震发生的次数符合上述 3 个假设, 并且每周发生的次数为 2 次).

(a) 接下来 2 周内至少发生 3 次地震的概率;

(b) 从现在开始到下次地震的持续时间的概率分布.

解: (a) 由 (7.5) 式, 有

$$\begin{aligned} P\{N(2) \geqslant 3\} &= 1 - P\{N(2) = 0\} - P\{N(2) = 1\} - P\{N(2) = 2\} \\ &= 1 - \mathrm{e}^{-4} - 4\mathrm{e}^{-4} - \frac{4^2}{2}\mathrm{e}^{-4} = 1 - 13\mathrm{e}^{-4} \end{aligned}$$

(b) 令 X 表示"从现在开始到下次地震时间的时间间隔"(单位: 周). 因为 X 大于 t 的充要条件是接下来的时间 t 内不发生地震, 由 (7.5) 式得

$$P\{X > t\} = P\{N(t) = 0\} = \mathrm{e}^{-\lambda t}$$

因此, 随机变量 X 的分布函数 F 为

$$F(t) = P\{X \leqslant t\} = 1 - P\{X > t\} = 1 - \mathrm{e}^{-\lambda t} = 1 - \mathrm{e}^{-2t}$$ ■

计算泊松分布函数

如果 X 服从参数为 λ 的泊松分布, 则

$$\frac{P\{X = i+1\}}{P\{X = i\}} = \frac{\mathrm{e}^{-\lambda}\lambda^{i+1}/(i+1)!}{\mathrm{e}^{-\lambda}\lambda^i/i!} = \frac{\lambda}{i+1} \tag{7.6}$$

从 $P\{X = 0\} = \mathrm{e}^{-\lambda}$ 开始, 利用 (7.6) 式可以计算以下

$$P\{X = 1\} = \lambda P\{X = 0\}$$

$$P\{X = 2\} = \frac{\lambda}{2} P\{X = 1\}$$

$$\vdots$$

$$P\{X = i+1\} = \frac{\lambda}{i+1} P\{X = i\}$$

网站上有一个程序, 利用式 (7.6) 来计算泊松分布的有关概率.

例 7f

(a) X 服从泊松分布, 均值为 100, 计算 $P\{X \leqslant 90\}$.

(b) Y 服从泊松分布, 均值为 1000, 计算 $P\{Y \leqslant 1075\}$.

解: 利用网站上的程序可求得

(a) $P\{X \leqslant 90\} \approx 0.1714$.

(b) $P\{Y \leqslant 1075\} \approx 0.9894$. ■

4.8 其他离散型分布

4.8.1 几何随机变量

考虑独立重复试验, 每次成功率为 $p, 0 < p < 1$, 一直进行直到试验成功. 如果令 X 表示需要试验的次数, 那么

$$P\{X = n\} = (1-p)^{n-1}p \qquad n = 1, 2, \cdots \tag{8.1}$$

上式成立是因为要使得 X 等于 n, 充分且必要条件是前 $n-1$ 次试验失败而第 n 次试验成功. 又因为假定各次试验都是相互独立的, 因此 (8.1) 式成立.

由于

$$\sum_{n=1}^{\infty} P\{X = n\} = p\sum_{n=1}^{\infty}(1-p)^{n-1} = \frac{p}{1-(1-p)} = 1$$

这说明试验最终会出现成功的概率为 1. 若随机变量的分布列由 (8.1) 式给出, 则称该随机变量为参数为 p 的几何(geometric)随机变量.

例 8a 一个坛子里有 N 个白球和 M 个黑球. 每次从中取出一个球, 观察球的颜色并放回, 重复这个过程, 直到取出一个黑球, 求以下事件概率:

(a) 正好取球 n 次; (b) 至少取球 k 次.

解: 令 X 表示要取出一个黑球需要取球的次数, 则 X 满足公式 (8.1), 其中 $p = M/(M+N)$, 因此

(a)

$$P\{X = n\} = \left(\frac{N}{M+N}\right)^{n-1}\frac{M}{M+N} = \frac{MN^{n-1}}{(M+N)^n}$$

(b)

$$\begin{aligned}
P\{X \geqslant k\} &= \frac{M}{M+N}\sum_{n=k}^{\infty}\left(\frac{N}{M+N}\right)^{n-1} \\
&= \left(\frac{M}{M+N}\right)\left(\frac{N}{M+N}\right)^{k-1}\Big/\left[1 - \frac{N}{M+N}\right] = \left(\frac{N}{M+N}\right)^{k-1}
\end{aligned}$$

问题 (b) 的答案可以直接得到, 因为至少需要 k 次取球意味前 $k-1$ 次拿到的都是白球, 亦即前 $k-1$ 次试验都失败. 这样, 对于一个服从几何分布的随机变量,

$$P\{X \geqslant k\} = (1-p)^{k-1} = \left(\frac{N}{M+N}\right)^{k-1}$$ ■

例 8b 计算几何随机变量的期望.

解: 令 $q = 1-p$, 我们有

$$\begin{aligned} E[X] &= \sum_{i=1}^{\infty} iq^{i-1}p = \sum_{i=1}^{\infty}(i-1+1)q^{i-1}p = \sum_{i=1}^{\infty}(i-1)q^{i-1}p + \sum_{i=1}^{\infty} q^{i-1}p \\ &= \sum_{j=0}^{\infty} jq^j p + 1 = q\sum_{j=1}^{\infty} jq^{j-1}p + 1 = qE[X] + 1 \end{aligned}$$

因此 $pE[X] = 1$, 由此得到 $E[X] = 1/p$. 也就是说, 一个成功的概率为 p 的试验, 如果独立重复进行直到试验成功, 那么需要进行的试验的次数的期望等于 $1/p$. 举例来说, 掷一枚均匀骰子, 直到出现一次点数为 1, 需要掷的次数的期望为 6. ■

例 8c 计算几何随机变量的方差.

解: 为了计算 $\text{Var}(X)$, 先计算 $E[X^2]$, 记 $q = 1-p$, 我们有

$$\begin{aligned} E[X^2] &= \sum_{i=1}^{\infty} i^2 q^{i-1}p = \sum_{i=1}^{\infty}(i-1+1)^2 q^{i-1}p \\ &= \sum_{i=1}^{\infty}(i-1)^2 q^{i-1}p + \sum_{i=1}^{\infty} 2(i-1)q^{i-1}p + \sum_{i=1}^{\infty} q^{i-1}p = \sum_{j=0}^{\infty} j^2 q^j p + 2\sum_{j=1}^{\infty} jq^j p + 1 \\ &= qE[X^2] + 2qE[X] + 1 \end{aligned}$$

利用公式 $E[X] = 1/p$, 由上面的公式得到

$$pE[X^2] = \frac{2q}{p} + 1$$

因此,

$$E[X^2] = \frac{2q+p}{p^2} = \frac{q+1}{p^2}$$

再利用方差的公式得到

$$\text{Var}(X) = E[X^2] - (E[X])^2 = \frac{q+1}{p^2} - \frac{1}{p^2} = \frac{q}{p^2} = \frac{1-p}{p^2}$$ ■

4.8.2 负二项分布

考虑独立重复试验, 每次成功的概率为 $p, 0 < p < 1$, 试验一直进行到一共累计成功了 r 次为止. 令 X 表示此时试验的总次数, 则

$$P\{X = n\} = \binom{n-1}{r-1} p^r (1-p)^{n-r} \qquad n = r, r+1, \cdots \tag{8.2}$$

公式 (8.2) 之所以成立是因为, 要使得第 n 次试验时正好是第 r 次成功, 那么前 $n-1$ 次试验中有 $r-1$ 次成功, 且第 n 次试验必然是成功, "前 $n-1$ 次试验中有 $r-1$ 次成功" 这一事件的概率为

$$\binom{n-1}{r-1}p^{r-1}(1-p)^{n-r}$$

而 "第 n 次试验成功" 的概率为 p. 因为这两事件相互独立, 将两个概率值相乘就得到 (8.2). 要证明如果试验一直进行, 那么最终一定能得到 r 次成功, 只需用分析的方法证明

$$\sum_{n=r}^{\infty} P\{X=n\} = \sum_{n=r}^{\infty} \binom{n-1}{r-1}p^{r}(1-p)^{n-r} = 1 \tag{8.3}$$

或者我们可给出如下的概率论证的方法: 得到 r 次成功所需的试验次数可以分解为 $Y_1+Y_2+\cdots+Y_r$, 其中 Y_1 表示第一次成功时试验的次数, Y_2 表示第一次成功之后, 直到第二次成功时所需的试验次数, Y_3 表示第二次成功之后, 直到第三次成功所需的试验次数, 等等. 因为试验是相互独立的, 且每次成功的概率都为 p, 所以, $Y_1, Y_2, \cdots, Y_r$ 都为几何随机变量. 而几何随机变量 Y_i 都是以概率 1 取有限值, 因此, $\sum_{i=1}^{r} Y_i$ 一定为有限值. 这样 (8.3) 式就得到了证明.

若随机变量 X 的分布列由 (8.3) 式给出, 那么称 X 为参数为 (r,p) 的**负二项** (negative binomial) **随机变量**. 注意几何随机变量就是参数为 $(1,p)$ 的负二项随机变量.

接下来的例子, 我们将要利用负二项分布来得到关于点的问题的另一个解法.

例 8d 进行独立重复试验, 每次成功的概率为 p, 求第 r 次成功发生在 m 次失败之前的概率.

解: 注意到当且仅当第 r 次成功的时刻不晚于 $r+m-1$, 才能保证在 m 次失败之前出现第 r 次成功. 这是因为, 如果在 $r+m-1$ 次试验之前或此时已经有 r 次成功发生, 那么在 m 次失败之前必然有 r 次成功, 反之也成立. 因此, 利用公式 (8.2), 所求概率为

$$\sum_{n=r}^{r+m-1} \binom{n-1}{r-1}p^{r}(1-p)^{n-r}$$ ■

例 8e (巴拿赫火柴问题) 某个抽烟的数学家一直就随身带着两盒火柴, 一盒放在左边口袋, 另一盒放在右边口袋. 每次他需要火柴时, 都是随机地从两个口袋中任取一盒, 并取出其中一根. 假设开始时两盒中都是 N 根火柴, 问在他第一次发现其中有一个盒子已经空了的时候, 另一盒中恰好有 k 根火柴的概率有多大? $k=0,1,2,\cdots,N$.

解: 令 E 表示事件 "数学家第一次发现右边口袋里的火柴盒是空的, 而此时左边口袋里的火柴盒里还有 k 根火柴", 这个事件发生当且仅当第 $(N+1+N-k)$

次抽取根火柴时候正好取中的是右边口袋, 而且是第 $(N+1)$ 次取中右边口袋, 因此, 利用公式 (8.2)($p=1/2, r=N+1, n=2N-k+1$), 有

$$P(E)=\binom{2N-k}{N}\left(\frac{1}{2}\right)^{2N-k+1}$$

另外, 还有同样概率的事件是第一次发现左边口袋里的火柴盒是空的, 而此时右边口袋火柴盒里恰好还有 k 根火柴. 而这两个事件又是互不相容的, 因此所求概率为

$$2P(E)=\binom{2N-k}{N}\left(\frac{1}{2}\right)^{2N-k}$$ ■

例 8f 计算参数为 (r, p) 的负二项随机变量的期望值和方差.

解:

$$\begin{aligned}
E[X^k] &= \sum_{n=r}^{\infty} n^k \binom{n-1}{r-1} p^r (1-p)^{n-r} \\
&= \frac{r}{p}\sum_{n=r}^{\infty} n^{k-1}\binom{n}{r} p^{r+1}(1-p)^{n-r} \qquad \text{因为} n\binom{n-1}{r-1}=r\binom{n}{r} \\
&= \frac{r}{p}\sum_{m=r+1}^{\infty}(m-1)^{k-1}\binom{m-1}{r}p^{r+1}(1-p)^{m-(r+1)} \qquad \text{令} m=n+1 \\
&= \frac{r}{p}E[(Y-1)^{k-1}]
\end{aligned}$$

其中 Y 是一个参数为 $(r+1,p)$ 的负二项随机变量. 在上式中令 $k=1$, 可以得到

$$E[X]=\frac{r}{p}$$

在上式中令 $k=2$, 并利用负二项分布的随机变量的期望的公式, 可以得到

$$E[X^2]=\frac{r}{p}E[Y-1]=\frac{r}{p}\left(\frac{r+1}{p}-1\right)$$

因此,

$$\mathrm{Var}(X)=\frac{r}{p}\left(\frac{r+1}{p}-1\right)-\left(\frac{r}{p}\right)^2=\frac{r(1-p)}{p^2}$$ ■

从例 8f 可以看出: 如果进行独立重复试验, 每次成功的概率为 p, 则需要累积 r 次成功的总试验次数的期望值和方差分别为 r/p 和 $r(1-p)/p^2$.

由于几何随机变量就是参数 $r=1$ 的负二项随机变量, 由上面的例子可得参数为 p 的几何随机变量的方差为 $(1-p)/p^2$, 这样就验证了例 8c 的结果.

例 8g 连续掷一枚骰子, 一直到点数 1 出现了 4 次, 求投掷总次数的期望值和方差.

解: 因为我们所关心的随机变量 X(投掷总次数) 服从负二项分布, 参数 $r=4$, $p=1/6$. 因此

$$E[X]=24 \qquad \mathrm{Var}(X)=\frac{4\times\frac{5}{6}}{\left(\frac{1}{6}\right)^2}=120$$ ■

4.8.3 超几何随机变量

一个坛子里有 N 个球, 其中 m 个白球, $N-m$ 个黑球, 随机地 (无放回) 从中取出大小为 n 的样本, 令 X 表示取出来的白球数, 那么

$$P\{X=i\}=\frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}} \qquad i=0,1,\cdots,n \tag{8.4}$$

一个随机变量 X, 如果其概率分布由公式 (8.4) 给出, 那么 X 就称为*超几何* (hypergeometirc) *随机变量*.

注释 虽然我们把超几何随机变量取值的概率从 0 写到了 n, 但事实上 $P\{X=i\}$ 等于 0, 除非 i 满足 $n-(N-m)\leqslant i\leqslant \min(n,m)$. 然而, 式 (8.4) 一般来说是成立的, 因为为了方便, 我们规定了在 $k<0$ 或 $r<k$ 时, $\binom{r}{k}$ 等于 0. ■

例 8h 栖居于某地区的动物个数 N 是未知的, 为了得到对 N 的大致估计, 生态学家们常常进行如下的试验. 他们先在这个地区捕捉一些动物, 比如说 m 个, 然后标上记号放掉它们. 过一段时间, 当这些标有记号的动物充分散布到整个地区后, 再捉一批, 比如说 n 个. 设 X 为第二批捉住的 n 个动物中标过记号的动物个数. 如果假设两次捕捉期间动物的总数没有发生变化, 而且捉住每一只动物的可能性是一样的, 那么 X 为一超几何随机变量, 满足

$$P\{X=i\}=\frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}\equiv P_i(N)$$

其中 i 为 X 的观测值. 由于 $P_i(N)$ 是这个地区事实上总共有 N 个动物的条件下观测事件 X 取值的概率, 故使 $P_i(N)$ 达到最大值的 N 值应当是动物总数 N 的合理的估计. 这样的估计称为*极大似然* (maximum likelihood) *估计*. 更多的这种估计的例子可参见理论习题 13 和理论习题 18.

求 $P_i(N)$ 最大值的最简单的方法是: 首先注意到

$$\frac{P_i(N)}{P_i(N-1)}=\frac{(N-m)(N-n)}{N(N-m-n+i)}$$

要使上述比值大于 1, 当且仅当

$$(N-m)(N-n) \geqslant N(N-m-n+i)$$

或等价地, 当且仅当

$$N \leqslant \frac{mn}{i}$$

可见, $P_i(N)$ 是先上升, 然后下降, 且在不超过 mn/i 的最大整数处达到其最大值. 这个最大整数就是 N 的极大似然估计. 例如, 假定第一次捕捉到了 $m=50$ 只动物, 标上记号后放掉. 第二次又捕捉了 $n=40$ 只动物, 其中标有记号的有 $i=4$ 只, 那么, 我们就可估计出这地区大约有 500 只动物. (要注意上述估计还可以这样求得: 在这个地区内, 标有记号的动物所占的比例为 m/N, 应当近似地等于第二次捕捉的动物中做过标记的动物所占的比例 i/n.) ■

例 8i 某采购员购买一种 10 个一包的电子元件. 从一包中随机地抽查 3 个, 如果这 3 个元件都是好的, 才买下这一包. 假定含有 4 个残次元件的包数占 30%, 而其余 70% 每包只有一个残次元件. 试问被这个采购员拒绝的包数占多大比例?

解: 设 A 表示"采购员买下某一包"这一事件, 则

$$\begin{aligned}P(A) &= P(A|\text{这包有 4 个残次品}) \times \frac{3}{10} + P(A|\text{这包有 1 个残次品}) \times \frac{7}{10} \\ &= \frac{\binom{4}{0}\binom{6}{3}}{\binom{10}{3}} \times \frac{3}{10} + \frac{\binom{1}{0}\binom{9}{3}}{\binom{10}{3}} \times \frac{7}{10} = \frac{54}{100}\end{aligned}$$

从而, 将有 46% 的包被采购员拒绝. ■

从 N 个球 (白球比例为 $p=m/N$) 里, 无放回随机抽取 n 个球, 那么取中的白球数为超几何随机变量. 如果对于 n 来说, m 和 N 很大的话, 那么有放回和无放回抽球没什么差别. 因为当 m 和 N 很大时, 不管前面抽了多少球, 接下来的抽到的是白球的概率仍然近似等于 p . 也就是说, 直觉认为, 当 m 和 N 相比 n 很大时, X 的分布列应该近似等于参数为 (n,p) 的二项随机变量的分布列. 为了证明这个直觉, 注意到如果 X 为一超几何随机变量, 那么对 $i \leqslant n$, 有

$$\begin{aligned}P\{X=i\} &= \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}} \\ &= \frac{m!}{(m-i)!\,i!} \cdot \frac{(N-m)!}{(N-m-n+i)!\,(n-i)!} \cdot \frac{(N-n)!\,n!}{N!} \\ &= \binom{n}{i}\frac{m}{N} \cdot \frac{m-1}{N-1} \cdots \frac{m-i+1}{N-i+1} \cdot \frac{N-m}{N-i} \cdot \frac{N-m-1}{N-i-1} \cdots \frac{N-m-(n-i-1)}{N-i-(n-i-1)}\end{aligned}$$

$$\approx \binom{n}{i}p^i(1-p)^{n-i}$$

其中, 最后等式成立的条件是 $p=m/N$ 且 m 和 N 相对 n 和 i 来说都很大.

例 8j 试计算服从参数为 (n,N,m) 的超几何随机变量 X 的期望和方差.

解:

$$E[X^k]=\sum_{i=0}^{n}i^kP\{X=i\}=\sum_{i=1}^{n}i^k\binom{m}{i}\binom{N-m}{n-i}\Big/\binom{N}{n}$$

利用恒等式

$$i\binom{m}{i}=m\binom{m-1}{i-1}\qquad 和\qquad n\binom{N}{n}=N\binom{N-1}{n-1}$$

可以得到

$$\begin{aligned}E[X^k]&=\frac{nm}{N}\sum_{i=1}^{n}i^{k-1}\binom{m-1}{i-1}\binom{N-m}{n-i}\Big/\binom{N-1}{n-1}\\&=\frac{nm}{N}\sum_{j=0}^{n-1}(j+1)^{k-1}\binom{m-1}{j}\binom{N-m}{n-1-j}\Big/\binom{N-1}{n-1}=\frac{nm}{N}E[(Y+1)^{k-1}]\end{aligned}$$

其中, Y 是一个服从超几何分布的随机变量, 其参数为 $(n-1,N-1,m-1)$. 因此, 在上面的等式中令 $k=1$, 有 $E[X]=nm/N$, 换言之, 如果从 N 个球 (其中 m 个白球) 中随机抽取 n 个, 那么其中白球数的期望为 nm/N.

在上面的式子中令 $k=2$, 可以得到

$$E[X^2]=\frac{nm}{N}E[Y+1]=\frac{nm}{N}\left[\frac{(n-1)(m-1)}{N-1}+1\right]$$

后一个等式用到了前面关于服从超几何分布的随机变量 Y 的期望的计算结果. 又由 $E[X]=nm/N$, 我们可以推导出

$$\mathrm{Var}(X)=\frac{nm}{N}\left[\frac{(n-1)(m-1)}{N-1}+1-\frac{nm}{N}\right]$$

令 $p=m/N$, 且利用等式

$$\frac{m-1}{N-1}=\frac{Np-1}{N-1}=p-\frac{1-p}{N-1}$$

得到

$$\mathrm{Var}(X)=np[(n-1)p-(n-1)\frac{1-p}{N-1}+1-np]=np(1-p)\left(1-\frac{n-1}{N-1}\right)\qquad\blacksquare$$

注释 例 8j 已经指出, 从 N 个球 (白球的比例为 p) 里随机无放回抽取 n 个球, 那么抽取到的白球数的期望为 np. 而且, 当 N 相对 n 很大 (这样 $(N-n)/(N-1)$ 近似等于 1) 时, 有 $\text{Var}(X) \approx np(1-p)$, 也就是说, $E[X]$ 与有放回抽球情形下 (此时白球数为参数为 (n, p) 的二项随机变量) 是一样的. 而且, 如果总的球数很大, 那么 $\text{Var}(X)$ 近似等于有放回的情形. 当然, 这正是我们之前的猜测: 当坛子里的球数很大时, 抽取的白球数近似具有二项随机变量的分布列. ■

4.8.4 ζ (Zipf) 分布

一个随机变量称为服从 ζ 分布(有时也称为 Zipf 分布), 如果其分布列如下:

$$P\{X = k\} = \frac{C}{k^{\alpha+1}} \qquad k = 1, 2, \cdots$$

其中 $\alpha > 0$ 为参数. 因为概率之和必然等于 1, 因此有

$$C = \left[\sum_{k=1}^{\infty} \left(\frac{1}{k}\right)^{\alpha+1}\right]^{-1}$$

ζ 分布的名字来源于以下函数

$$\zeta(s) = 1 + \left(\frac{1}{2}\right)^s + \left(\frac{1}{3}\right)^s + \cdots + \left(\frac{1}{k}\right)^s + \cdots$$

它是数学中熟知的黎曼 ζ 函数 (起源于德国数学家 G. F. B. 黎曼).

ζ 分布曾被意大利经济学家 Pareto 用来描述某个给定国家的家庭收入的分布. 然而, 把这一分布运用到更广泛的各种不同的领域, 从而推广其应用的是 G. K. Zipf, 因此又叫 Zipf 分布.

4.9 随机变量和的期望值

随机变量的一个十分重要的性质是: 诸随机变量之和的期望值等于它们的期望值的和. 本节中我们将证明这个性质, 不过我们对样本空间作一个限制, 那就是样本空间 S 是有限或可数无限集合. 当然, 没有这个限制, 随机变量还是具有这个重要性质的 (在理论习题中将给出证明的纲要). 我们作这样的限制, 不但可使证明变得简要而且可使期望变得更加直观. 因此, 在本节中, 我们将假定样本空间 S 为有限或可数无限集合.

对于随机变量 X, 用 $X(s)$ 表示当试验结果为 $s \in S$ 时随机变量的取值. 当 X, Y 是随机变量, 则它们的和 $Z = X + Y$ 也是随机变量, 并且 $Z(s) = X(s) + Y(s)$.

例 9a 设试验为抛掷一枚硬币共 5 次, 此时试验结果是正面与反面的一个序列 (长度为 5). 设 X 为前面三次抛掷得到正面向上的次数, Y 为后面两次抛掷所

得到的正面向上的次数. 令 $Z = X + Y$, 例如, $S = (\mathrm{h}, \mathrm{t}, \mathrm{h}, \mathrm{t}, \mathrm{h})$(其中 h 表示正面向上, t 表示反面向上),

$$X(s) = 2 \quad Y(s) = 1 \quad Z(s) = X(s) + Y(s) = 3$$

其意义为: 结果 $S = (\mathrm{h}, \mathrm{t}, \mathrm{h}, \mathrm{t}, \mathrm{h})$ 中前 3 次抛掷得到 2 次正面向上, 后 2 次抛掷得到 1 次正面向上, 而 $Z(s)$ 刚好是 5 次抛掷得到正面向上的总次数 (3 次). ■

记 $p(s) = P(\{s\})$ 表示试验结果 s 的概率. 对于任何事件 A, 可以写成有限个或可数无限个互不相容的事件 $\{s\}$ 之和, 利用概率公理可知

$$P(A) = \sum_{s \in A} p(s)$$

当 $A = S$ 时上式变成

$$1 = \sum_{s \in S} p(s)$$

现在考虑随机变量 X 和它的期望 $E[X]$. 由于 X 的取值为 $X(s)$, 其中 s 为试验结果, $E[X]$ 在直观上可以认为是 $X(s)$ 的加权平均, 其权值刚好是 s 出现的概率 $p(s)$. 现在证明这个直观的结论.

命题 9.1

$$E[X] = \sum_{s \in S} X(s)p(s)$$

证明: 设 X 的取值范围为 $x_i, i \geqslant 1$. 对每一个 i, 令 S_i 表示事件 $\{X = x_i\}$, 即 $S_i = \{s : X(s) = x_i\}$. 由期望的定义, 通过一串等式就可以得到命题的结论,

$$\begin{aligned}
E[X] &= \sum_i x_i P\{X = x_i\} \\
&= \sum_i x_i P(S_i) = \sum_i x_i \sum_{s \in S_i} P(s) \\
&= \sum_i \sum_{s \in S_i} x_i p(s) = \sum_i \sum_{s \in S_i} X(s)p(s) \\
&= \sum_{s \in S} X(s)p(s)
\end{aligned}$$

其中第一个等式是随机变量期望的定义, 最后的等式是样本空间 S 为诸不相容的事件 $S_1, S_2, \cdots$ 之并. ■

例 9b 将一个硬币独立地抛掷两次, 假定正面向上的概率为 p, 令 X 表示正面向上的次数. 由于

$$\begin{aligned}
P(X = 0) &= P(t, t) = (1 - p)^2 \\
P(X = 1) &= P(h, t) + P(t, h) = 2p(1 - p) \\
P(X = 2) &= P(h, h) = p^2
\end{aligned}$$

由期望的定义知

$$E[X] = 0 \cdot (1-p)^2 + 1 \cdot 2p(1-p) + 2p^2 = 2p$$

而由命题 9.1 得到

$$\begin{aligned} E[X] &= X(h,h)p^2 + X(h,t)p(1-p) + X(t,h)(1-p)p + X(t,t)(1-p)^2 \\ &= 2p^2 + p(1-p) + (1-p)p + 0 = 2p \end{aligned}$$

两种计算方法得到的结果相同. ■

现在我们来证明关于随机变量和的期望的一个十分重要且有用的性质.

推论 9.2　对于随机变量 $X_1, \cdots, X_n$, 下式成立

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$$

证明: 令 $Z = \sum_{i=1}^{n} X_i$, 利用命题 9.1 得

$$\begin{aligned} E[Z] &= \sum_{s \in S} Z(s)p(s) \\ &= \sum_{s \in S} \Big(X_1(s) + X_2(s) + \cdots + X_n(s)\Big)p(s) \\ &= \sum_{s \in S} X_1(s)p(s) + \sum_{s \in S} X_2(s)p(s) + \cdots + \sum_{s \in S} X_n(s)p(s) \\ &= E[X_1] + E[X_2] + \cdots + E[X_n] \end{aligned}$$

□

例 9c　设抛掷 n 颗均匀的骰子, 求得到点数之总和的期望值.

解: 记 X 为得到点数之总和, 利用下列 X 的表达式来计算 $E[X]$:

$$X = \sum_{i=1}^{n} X_i$$

其中 X_i 为第 i 颗骰子出现的点数. 由于每颗骰子都是均匀的, 得到

$$E[X_i] = \sum_{j=1}^{6} j(1/6) = 21/6 = 7/2$$

这样,

$$E[X] = E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i] = 3.5n$$

例 9d　现设一共有 n 个试验, 其中第 i 次试验成功的概率为 $p_i, i = 1, 2, \cdots, n$. 求成功次数的期望值.

解: 令

$$X_i = \begin{cases} 1, & \text{试验 } i \text{ 成功} \\ 0, & \text{试验 } i \text{ 失败} \end{cases}$$

n 次试验成功次数 X 具有下面的表达式

$$X = \sum_{i=1}^{n} X_i$$

利用推论 9.2 得到

$$E[X] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} p_i$$

注意, 这个结果不要求各次试验相互独立. 而二项试验是 n 次独立重复试验, 每次试验成功概率为 $p_i = p$, 这是本例的一种特殊情况. 因此, 利用本例的结论可得, n 次试验中成功次数的期望值为 np. 我们也可以将超几何随机变量进行分解. 设坛子中一共有 N 个球, 其中有 m 个白球. 现在从坛子中无放回地抽取几个球. 坛子中的任意一个球 (N个) 都有相同的机会在第 i 次抽中, 因此第 i 次抽取抽中白球的概率为 $p = m/N$, 尽管各次抽取互相不独立, 但是无放回抽样还是符合本例的要求, 抽得白球数的期望为 $np = nm/N$. ■

例 9e 在例 9d 中讨论了两种特别的情况, 二项随机变量和超几何随机变量, 试导出这两个随机变量的方差.

解: 令 X 表示 n 次试验中成功的次数 (在超几何随机变量的情况, 第 i 次取出白球视为该次试验成功). 如上例所示, X 可以写成 $X = \sum_{i=1}^{n} X_i$. 我们得到

$$\begin{aligned} E[X^2] &= E\left[\left(\sum_{i=1}^{n} X_i\right)\left(\sum_{j=1}^{n} X_j\right)\right] = E\left[\sum_{i=1}^{n} X_i\left(X_i + \sum_{j\neq i} X_j\right)\right] \\ &= E\left[\sum_{i=1}^{n} X_i^2 + \sum_{i=1}^{n}\sum_{j\neq i} X_i X_j\right] = \sum_{i=1}^{n} E[X_i^2] + \sum_{i=1}^{n}\sum_{j\neq i} E[X_i X_j] \qquad (9.1) \\ &= \sum_{i} p_i + \sum_{i=1}^{n}\sum_{j\neq i} E[X_i X_j] \end{aligned}$$

上式中最后一个等式用到了 $X_i^2 = X_i$. 由于 X_i, X_j 只取 0, 1 两个值, 我们得到

$$X_i X_j = \begin{cases} 1 & X_i = 1, X_j = 1 \\ 0 & \text{其他} \end{cases}$$

从而

$$E[X_i X_j] = P\{X_i = 1, X_j = 1\} = P(\text{第 } i \text{ 次试验和第 } j \text{ 次试验都成功})$$

当 X 是二项随机变量时, X_i 与 X_j 相互独立, 这样

$$E[X_iX_j] = p^2, \quad i \neq j$$

再利用 (9.1) 式, 可得到

$$E[X^2] = np + n(n-1)p^2$$

从而

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2 = np + n(n-1)p^2 - n^2p^2 = np(1-p)$$

当 X 是超几何随机变量的时候, $P\{X_i = 1, X_j = 1\}$ 可以这样求得: 在第 i 次取得白球情况下, 第 j 次抽取时, 剩下的 $N-1$ 个球具有相同的机会被抽取, 而抽到白球的条件概率为 $(m-1)/(N-1)$. 这样

$$P\{X_i = 1, X_j = 1\} = P\{X_i = 1\}P\{X_j = 1|X_i = 1\} = \frac{m}{N} \cdot \frac{m-1}{N-1}$$

利用 (9.1) 式 ($p_i = m/N$) 可得

$$E[X^2] = \frac{nm}{N} + n(n-1)\frac{m}{N}\frac{m-1}{N-1}$$

其方差为

$$\mathrm{Var}(X) = \frac{nm}{N} + n(n-1)\frac{m}{N}\frac{m-1}{N-1} - \left(\frac{nm}{N}\right)^2$$

如例 8j 一样, 上式可化简为

$$\mathrm{Var}(X) = np(1-p)\left(1 - \frac{n-1}{N-1}\right)$$

其中 $p = m/N$ ■

4.10 分布函数的性质

回顾 X 的分布函数 F 之定义, $F(b)$ 表示随机变量取值小于或等于 b 的概率. 以下是一些有关分布函数的性质.

1. F 是一个非降函数, 也即, 如果 $a < b$, 那么 $F(a) \leqslant F(b)$;

2. $\lim\limits_{b \to \infty} F(b) = 1$;

3. $\lim\limits_{b \to -\infty} F(b) = 0$;

4. F 是右连续的, 也即, 对于任一 b 和一个递减且收敛于 b 的序列 $b_n, n \geqslant 1$, 有 $\lim\limits_{n \to \infty} F(b_n) = F(b)$.

性质 1 成立的原因是注意到在 4.1 节中已经指出, 因为对 $a < b$, 事件 $\{X \leqslant a\}$ 包含在事件 $\{X \leqslant b\}$ 中, 因此, 前者的概率不可能比后者大. 性质 2, 3, 4 成立都因

为概率的连续性 (2.6 节). 例如, 为了证明性质 2, 注意到, 如果 b_n 递增到 ∞, 那么事件序列 $\{X \leqslant b_n\}, n \geqslant 1$ 为递增事件列, 它们的并为事件 $\{X < \infty\}$. 因此, 利用概率的连续性, 有 $\lim\limits_{n\to\infty} P\{X \leqslant b_n\} = P\{X < \infty\} = 1$, 这样, 性质 2 就得到了证明.

性质 3 的证明类似, 留作习题. 为了证明性质 4, 注意到, 如果 b_n 递减到 b, 那么 $\{X \leqslant b_n\}, n \geqslant 1$, 为递减事件序列, 它们的交为 $\{X \leqslant b\}$. 因此, 根据概率的连续性, 可得 $\lim\limits_{n\to\infty} P\{X \leqslant b_n\} = P\{X \leqslant b\}$, 这样就证明了性质 4.

所有有关 X 的概率都可以通过其分布函数 F 进行计算. 例如

$$P\{a < X \leqslant b\} = F(b) - F(a) \qquad \text{对任意} a < b \tag{10.1}$$

将事件 $\{X \leqslant b\}$ 写成互不相容事件 $\{X \leqslant a\}$ 和 $\{a < X \leqslant b\}$ 的并也可以得到这点, 即 $\{X \leqslant b\} = \{X \leqslant a\} \cup \{a < X \leqslant b\}$, 因此, $P\{X \leqslant b\} = P\{X \leqslant a\} + P\{a < X \leqslant b\}$. 这样, 公式 (10.1) 就成立了.

如果要计算"X 严格小于 b"的概率, 那么再次利用概率的连续性可以得到

$$P\{X < b\} = P\left(\lim_{n\to\infty}\left\{X \leqslant b - \frac{1}{n}\right\}\right) = \lim_{n\to\infty} P\left(X \leqslant b - \frac{1}{n}\right) = \lim_{n\to\infty} F\left(b - \frac{1}{n}\right)$$

注意 $P\{X < b\}$ 并不一定等于 $F(b)$, 因为定义 $F(b)$ 事件为 $\{x < b\} + \{x = b\}$.

例 10a 随机变量 X 的概率分布函数如下:

$$F(x) = \begin{cases} 0 & x < 0 \\ x/2 & 0 \leqslant x < 1 \\ 2/3 & 1 \leqslant x < 2 \\ 11/12 & 2 \leqslant x < 3 \\ 1 & 3 \leqslant x \end{cases}$$

$F(x)$ 的图像见图 4.8, 计算:

(a) $P\{X < 3\}$; (b) $P\{X = 1\}$; (c) $P\{X > 1/2\}$; (d) $P\{2 < X \leqslant 4\}$.

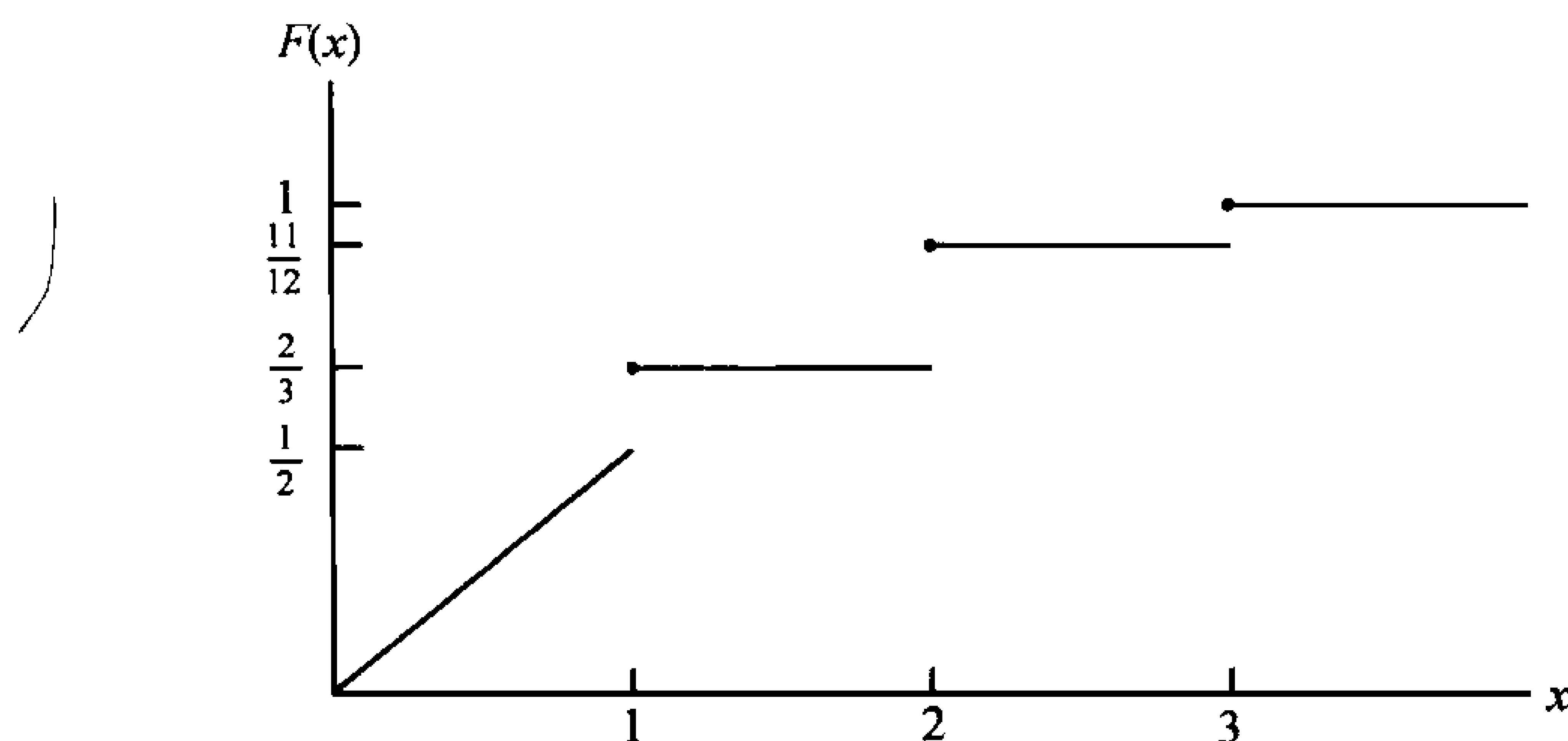


图 4.8 $F(x)$ 的示意图

解: (a) $P\{X<3\}=\lim\limits_{n\to\infty}P\left\{X\leqslant 3-\dfrac{1}{n}\right\}=\lim\limits_{n\to\infty}F\left(3-\dfrac{1}{n}\right)=\dfrac{11}{12}$

(b) $P\{X=1\}=P\{X\leqslant 1\}-P\{X<1\}$

$$=F(1)-\lim_{n\to\infty}F\left(1-\frac{1}{n}\right)=\frac{2}{3}-\frac{1}{2}=\frac{1}{6}$$

(c) $P\left\{X>\dfrac{1}{2}\right\}=1-P\left\{X\leqslant\dfrac{1}{2}\right\}=1-F\left(\dfrac{1}{2}\right)=\dfrac{3}{4}$

(d) $P\{2<X\leqslant 4\}=F(4)-F(2)=\dfrac{1}{12}$ ■

小　结

定义在试验结果上的实值函数称为**随机变量**(random variable).

如果 X 是一随机变量, 那么如下定义的函数 $F(x)$

$$F(x)=P\{X\leqslant x\}$$

称为随机变量 X 的**分布函数**(distribution function). 任意有关 X 的概率都可以通过 F 进行计算.

若一个随机变量的可能取值的集合是有限集, 或者可数无限集, 那么称该随机变量为**离散型**随机变量. 如果 X 是一个离散型随机变量, 那么函数

$$p(x)=P\{X=x\}$$

称为 X 的概率分布列或分布列. 另外, 如下定义的 $E[X]$

$$E[X]=\sum_{x:p(x)>0}xp(x)$$

称为 X 的**期望值**(expected value), $E[X]$ 通常也称为 X 的**均值**(mean) 或**期望**(expectation).

设 $g(x)$ 是一个实值函数, 则对于离散型随机变量 X, 关于 $E[g(X)]$ 的一个有用的计算公式为:

$$E[g(X)]=\sum_{x:p(x)>0}g(x)p(x)$$

随机变量 X 的**方差**(variance), 记为 $\mathrm{Var}(X)$, 定义如下:

$$\mathrm{Var}(X)=E[(X-E[X])^2]$$

方差等于 X 与它的期望的差的平方的期望, 它度量了 X 可能取值的分散程度. 下面是一个有用的恒等式

$$\mathrm{Var}(X)=E[X^2]-(E[X])^2$$

$\sqrt{\mathrm{Var}(X)}$ 称为 X 的**标准差**(standard deviation).

本章中, 我们介绍了一些常用的离散型随机变量.

若随机变量 X 的分布列为

$$p(i)=\binom{n}{i}p^i(1-p)^{n-i} \qquad i=0,\cdots,n$$

则 X 称为参数为 (n,p) 的二项随机变量. 该随机变量可解释为 n 次独立重复试验中试验成功的次数, 而每次试验成功的概率为 p. 它的均值和方差如下

$$E[X]=np \qquad \operatorname{Var}(X)=np(1-p)$$

若随机变量 X 的分布列为

$$p(i)=\frac{\mathrm{e}^{-\lambda}\lambda^i}{i!} \qquad i=0,1,\cdots$$

则 X 称为参数为 λ 的泊松随机变量. 如果进行多次 (近似) 独立的试验, 而且每次成功的概率都较小, 那么总的成功次数就近似为泊松随机变量. 泊松随机变量的均值和方差都等于其参数 λ, 也即 $E[X]=\operatorname{Var}(X)=\lambda$.

若随机变量 X 的分布列为

$$p(i)=p(1-p)^{i-1} \qquad i=1,2,\cdots$$

则 X 称为参数为 p 的几何(geometric) 随机变量. 在独立重复试验序列中, 从开始直到第一次成功为止的试验次数就是几何随机变量, 其分布参数 p 就是每次试验成功的概率. 其均值和方差分别为

$$E[X]=\frac{1}{p} \qquad \operatorname{Var}(X)=\frac{1-p}{p^2}$$

若随机变量 X 的分布列为

$$p(i)=\binom{i-1}{r-1}p^r(1-p)^{i-r} \qquad i=r,r+1,\cdots$$

则 X 称为参数为 (r,p) 的负二项 (negative binomial) 随机变量. 在独立重复试验序列中, 从开始直到第 r 次成功为止的试验次数就是负二项随机变量. 分布列中的参数 p 就是每次试验成功的概率. 其均值和方差分别为

$$E[X]=\frac{r}{p} \qquad \operatorname{Var}(X)=\frac{r(1-p)}{p^2}$$

设从一个装有 N 个球, 其中 m 个白球的坛子里, 随机抽取 n 个球, 其中白球的数目就是参数为 (n,N,m) 的超几何随机变量. 其分布列为

$$p(i)=\frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}} \qquad i=0,1,\cdots,m$$

其均值和方差分别为

$$E[X]=np \qquad \mathrm{Var}(X)=\frac{N-n}{N-1}np(1-p)$$

其中 $p=m/N$.

期望的一个重要性质是：随机变量和的期望等于它们的期望之和. 即

$$E\left[\sum_{i=1}^{n}X_i\right]=\sum_{i=1}^{n}E[X_i]$$

习 题

1. 坛子里有 8 个白球, 4 个黑球, 2 个橙色的球, 随机从中抽取 2 个. 假设抽取的球中每一个黑球能赢得 2 元, 每个白球要输掉 1 元. 令 X 表示最后赢得的数目, 那么 X 的可能取值是哪些? 取这些值的概率是多大?

2. 掷 2 枚均匀的骰子. 令 X 等于 2 枚骰子的点数乘积, 计算 $P\{X=i\}, i=1,2,\cdots,36$.

3. 掷 3 枚骰子. 假定 $6^3=216$ 种结果都是等可能的, 记 X 表示 3 枚骰子的点数之和, 计算 X 取各可能值的概率.

4. 5 个男生和 5 个女生依照他们的测验成绩排名. 假定没有两个学生的成绩是相同的, 而且所有 10! 种可能排名都是等可能的. 令 X 表示成绩最高的女生在全体同学中的排名 (比如, $X=1$ 表示第一名是女生). 求 $P\{X=i\}, i=1,2,3,\cdots,8,9,10$.

5. 掷一枚硬币 n 次, 令 X 表示得到的正面朝上数与反面朝上数之差. X 的可能取值是哪些?

6. 在习题 5 中, 如果硬币是均匀的, 对于 $n=3$, 计算 X 的分布列.

7. 掷一枚骰子 2 次, 以下随机变量的可能取值是哪些?

(a) 两次投掷出现的最大值;　　(b) 两次投掷出现的最小值;

(c) 两次投掷所出现的点数之和;　　(d) 第一次投掷的值减去第二次投掷的值.

8. 习题 7 中, 假设骰子是均匀的, 计算 (a) 到 (d) 里各随机变量可能取值的概率.

9. 在球是有放回的情形下重做例 1b.

10. 在例 1d 中, 计算在已知我们已经赢得钱的条件下, 共赢得 i 元的条件概率, $i=1,2,3$.

11. (a) 从 $\{1,2,\cdots,10^3\}$ 中随机选一个数 N, 且选中每个数的概率都一样. 问 N 能被 3 整除的概率是多大? 能被 5 整除呢? 能被 7 整除呢? 能被 15 整除呢? 能被 105 整除呢? 如果 10^3 换成 10^k, 并且 k 越来越大, 那么答案如何变化?

(b) 数论里有默比乌斯函数 $\mu(n)$, 它对所有正整数有定义：当 n 具有重复的素数因子时, 定义 $\mu(n)=0$. (例如, 当 $n=12$ 时, 其素因子为 $2\times2\times3$, 当 $n=49$ 时, $n=7\times7$, 对于这些 n,$\mu(n)=0$. 现设 N 为从 $\{1,2,\cdots,10^k\}$ 中随机取的一个数, 求 $\lim_{k\to\infty}P\{\mu(N)=0\}$.

提示：计算 $P\{\mu(N)\neq0\}$, 利用等式

$$\prod_{i=1}^{\infty}\frac{P_i^2-1}{P_i^2}=\frac{3}{4}\times\frac{8}{9}\times\frac{24}{25}\times\frac{48}{49}\cdots=\frac{6}{\pi^2}$$

其中 P_i 是第 i 小的素数 (1 不是素数).

12. 在“莫拉 (Morra) 二指”赌中, 两赌徒各伸出 1 或 2 个手指, 并同时猜对方伸出的手指数. 如果两人中只有一人猜对了, 那么他赢得两人伸出的手指数之和的钱 (单位: 元). 如果两人都猜对了或都没猜对, 则谁也不赢谁的钱. 现考虑某指定赌徒, 并设他在一局“莫拉二指”赌中赢得的钱数为 X.

(a) 如果两赌徒的行为是独立的, 并且假设每一赌徒将伸几个手指以及他猜测对方伸几个手指总共 4 种情况是等可能的. 试问 X 取哪些可能值? 取这些值的概率各是多少?

(b) 假定两赌徒的行为是独立的, 而且每个人猜对方要伸几个手指就决定自己伸几个手指, 且设每个赌徒伸 1 或 2 个手指是等可能的. 试问 X 取哪些可能值? 取这些值的概率各是多少?

13. 某销售员计划了两个销售会来推销百科全书. 第一个销售会成交的概率为 0.3, 第二个销售会成交的概率为 0.6, 且相互独立. 销售的百科全书有高级版, 价值 1000 元, 也有平装版, 价值 500 元, 销售任一种是等可能的. 计算总的销售值 X 的分布列.

14. 5 个不同的数随机分派给 1, 2, 3, 4, 5 号共 5 个人. 两人之间比较数的大小, 大者获胜. 最初, 1 号和 2 号比较, 胜者再同 3 号比较, 等等. 令 X 表示 1 号在比较中获胜的次数. 计算 $P\{X=i\}, i=0,1,2,3,4$.

15. 美国篮球联盟 (NBA) 选秀大会上包含有当年输赢记录最差的 11 支球队. 总共有 66 个球放入坛子里, 每个球都刻写了某个球队的名称. 有 11 个写了最差球队的名称, 有 10 个写了倒数第二球队的名称, 有 9 个写了倒数第三球队的名称, ……有 1 个写了倒数第 11 的球队的名称. 从中随机抽取一球, 其上面的球队获得第一轮选秀权, 它可以在补充球员时具有优先的选择权. 然后, 再抽取一个球, 如果它与第一次抽取的是不同的球队, 那么该球队获得第二轮选秀权. (如果与第一次抽取的是相同的球队, 那么放弃并重新抽球, 直到抽取到不同的球队为止.) 最后, 再抽取一个球 (假设抽取的球队与前两次不同), 其上球队获得第三轮选秀权. 剩下的第 4 轮到第 11 轮选秀权按照剩下球队的输赢顺序颠倒过来分配, 举例说, 如果记录最差的球队没有获得前三轮选秀权, 那么它将获得第四轮选秀权. 令 X 表示最差球队获得的选秀权轮数, 求 X 的分布列.

16. 在习题 15 中, 设 1 号球队为最差球队, 2 号球队为第二差球队, 等等. 令 Y_i 表示获得第 i 轮选秀权的球队号码. 因此, 如果第一轮选秀权属于 3 号球队, 那么 $Y_1=3$, 计算以下分布列: (a) Y_1; (b) Y_2; (c) Y_3.

17. 假设 X 的分布函数如下:

$$F(b)=\begin{cases}0 & b<0\\ b/4 & 0\leqslant b<1\\ 1/2+(b-1)/4 & 1\leqslant b<2\\ 11/12 & 2\leqslant b<3\\ 1 & 3\leqslant b\end{cases}$$

(a) 计算 $P\{X=i\}, i=1,2,3$. (b) 求 $P\left\{\frac{1}{2}<X<\frac{3}{2}\right\}$.

18. 独立重复投掷一枚均匀硬币 4 次, 令 X 表示“正面朝上的数目”, 画出随机变量 $X-2$ 的概率分布列.

19. 设 X 的分布函数由下式给出:

$$F(b)=\begin{cases}0 & b<0\\ 1/2 & 0\leqslant b<1\\ 3/5 & 1\leqslant b<2\\ 4/5 & 2\leqslant b<3\\ 9/10 & 3\leqslant b<3.5\\ 1 & b\geqslant 3.5\end{cases}$$

试求 X 的分布列.

20. 一本关于赌博的书中建议了如下轮盘赌"必胜策略". 赌徒押"红"1 元, 如果结果是"红"(概率为 18/38), 那他拿走 1 元利润, 并且离开. 如果赌徒输掉了这 1 元 (概率为 20/38), 他应该在下面两次轮盘赌中, 还押"红"1 元钱, 赌完两次以后就离开, 记 X 为这个人终止赌博时所赢的钱数,

(a) 求 $P\{X>0\}$. (b) 您相信该策略真是一个"必胜策略"吗? 给出您的解释.

(c) 求 $E[X]$.

21. 总共 4 辆大客车载着 148 名同学从同一个学校到足球馆, 车上分别有 40,33,25 和 50 名同学. 随机抽取一名同学, 令 X 表示该同学所在的车上的同学数. 同时随机选一位司机, 令 Y 表示他驾驶的车上的同学数.

(a) $E[X]$ 和 $E[Y]$ 哪个大? 为什么? (b) 计算 $E[X]$ 和 $E[Y]$.

22. 假设两个队进行一系列比赛, 一直到其中有一队赢了 i 局, 假设各局比赛胜负是相互独立的, 并且 A 队获胜概率为 p, 求出下列条件下比赛的局数的期望值:

(a) $i=2$. (b) $i=3$.

同时指出: 在两种情形下, 这个期望在 $p=1/2$ 时达到最大.

23. 假设你有 1000 元以及某商品, 该商品当前价格是每盎司 2 元. 假设一周后该商品的价格变成每盎司 1 元或者每盎司 4 元, 两种情况的可能性是一样的.

(a) 如果你的目的是使得一周后的期望财产达到最大, 你将采取什么策略?

(b) 如果你的目标是使得你拥有该商品的期望数一周后达到最大, 你又将采取什么策略?

24. A 和 B 进行如下赌博: A 写下 1 或 2 中的一个数, 而 B 要猜写的是哪一个, 如果 A 写下的数是 i 且 B 猜对了, 那么 B 从 A 那获得 i 元, 如果 B 猜错了, 那么 B 将付给 A 3/4 元. 如果 B 是随机地猜, 猜 1 的概率为 p, 猜 2 的概率为 $1-p$, 计算 B 赢钱的期望, 如果

(a) A 写的是 1; (b) A 写的是 2.

为使 B 的期望赢钱数的最小值达到最大, p 应该取什么值, 并求出这个极大的极小值. (B 的期望赢钱数不仅依赖于 p 的值, 还依赖于 A 的策略.)

现在考虑 A, 假设他也是随机地决定, 写 1 的概率为 q, 那么 A 的期望损失是多大? 如果

(c) B 猜的是 1; (d) B 猜的是 2.

当 q 取什么值时, 使得 A 的最大期望损失达到最小? 指出 A 的最小的最大期望损失刚好等于 B 的最大的最小期望赢钱数. 这个结果就是著名的极大极小定理, 它是由数学家约翰 · 冯诺依曼建立的, 它也是博弈论中的基本的数学结论. 这个公共值称为博弈者 B 的博弈值.

25. 抛掷两枚硬币, 其中一枚在抛掷时正面朝上的概率为 0.6, 另一枚正面向上的概率为 0.7. 假定两次抛掷是相互独立的, 令 X 为两次抛掷时正面向上的次数. (a) 求出 $P\{X = 1\}$; (b) 求 $E[X]$.

26. 一个人随机地从 $\{1, 2, \cdots, 10\}$ 中选择一个数, 然后让你猜这个数. 通过向对方问若干次以"是"或"否"为答案的问题, 你就可以根据推理确定这个数. 计算下列策略情况下, 需要提问次数的期望值.

(a) 第 i 个问题是:"是 i 吗? ", $i = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$.

(b) 每一个问题都尽可能去掉剩下数字的一半.[①]

27. 保险公司开出的保险单规定, 如果某个事件 E 在一年内发生了, 那么保险公司必须付出一笔钱 A. 如果保险公司估计事件 E 在一年内发生的概率为 p, 他应该向顾客收多少保险费, 他的期望收益才能达到 A 的 10%?

28. 一个箱子里有 20 件产品, 其中 4 件为残次品. 从中随机取 3 件作为样品, 计算样品中的残次品数量的期望值.

29. 某个机器停止工作有两个可能原因, 检查第一个可能原因需要花费 C_1 元, 经查如果确实是这个原因, 那么还需要修理费用 R_1 元才能修好机器. 类似地, 检查第二个可能原因需要花费 C_2 元, 如果确是这种原因需要修理费用为 R_2 元. 令 p 和 $1-p$ 分别表示机器停止工作由第一种和第二种原因引起的概率. 在 $p, C_i, R_i, i = 1, 2$ 满足何种条件下, 我们先检查第一种原因再检查第二种原因比反过来的次序能使得机器正常工作所需要的花费的期望值小?

注意: 如果第一次检查发现不是预期原因, 那么仍需检查第二种可能性.

30. 某人掷一枚均匀硬币直到第一次出现反面朝上. 如果反面朝上出现在第 n 次抛掷, 他将赢得 2^n 元, 令 X 表示他的赢利, 指出 $E[X] = +\infty$. 该问题就是圣彼得堡悖论.

(a) 你愿意付出 100 万元玩一局这个游戏吗?

(b) 如果每局付出 100 万, 你可以一直玩下去, 直到你想停止该游戏, 你会玩吗?

31. 每天晚上, 不同的气象工作者都将告诉我们明天下雨的概率. 为了判断他们的预测水平, 我们按如下方式给他们评分: 如果某个人说明天下雨的概率为 p, 那么他 (她) 的得分为

$$\begin{cases} 1 - (1-p)^2 & \text{如果第二天果然下雨} \\ 1 - p^2 & \text{否则} \end{cases}$$

我们将要在一个时期内记录这个分数, 并且推断平均分数最高者为最好的天气预测者. 现在假设某气象工作者也了解这点, 他 (她) 当然想使得其期望分数最大化. 如果他确信明天下雨的概率为 p^*, 那么他 (她) 应该给出怎样的 p 值才能使得期望分数最大化?

32. 100 人要做血液检查, 以便确定是否有某种疾病. 然而, 医院采用这样的办法: 将 100 人分成 10 人一组, 将 10 人的样本混在一起进行检查. 如果检验结果为阴性, 那么这一次测试对该组 10 人已经足够, 若混合样本为阳性, 那么还要对该组 10 个样本逐个进行检查, 这样就共要检查 11 次. 设每个人得病的概率为 0.1, 并且彼此相互独立. 计算对每一个 10 人小组检查的平均次数 (假定 10 人组内有一人得病, 混合血样就会显示阳性).

---

① 比如可以这样提问, 第一个问题: 这个数大于 5 吗? 如果答案是, 去掉了前面一半, 如果回答否, 去掉了后一半, 然后继续重复类似的问题. ——译者注

33. 某卖报小孩以 10 美分买进报纸并以 15 美分卖出, 然而, 不许他退还没有卖出的报纸. 如果卖的报纸的需求量是参数为 $n=10, p=1/3$ 的二项随机变量, 给出他大约应该进多少报纸以达到期望收益的最大化.

34. 在例 4b 中, 假设如果没有满足顾客的要求 (即顾客买商品时, 商店缺货), 也会导致每单位商品增加一个额外的成本 c(这常常称为信誉成本, 因为商店无法满足客户的要求), 计算该商店囤货 s 单位时其期望利润值, 并且计算能使得期望利润最大化的 s 值.

35. 盒子里有 5 个红弹子, 5 个蓝弹子, 随机从中取 2 个. 如果颜色相同, 那么赢得 1.10 元, 如果颜色不同, 将赢得 -1.00 元 (也即输掉 1 元), 计算

(a) 赢得的期望值. (b) 赢得的方差.

36. 考虑习题 22, 其中 $i=2$. 计算玩游戏局数的方差, 并证明当 $p=1/2$ 时, 该值取得最大值.

37. 计算习题 21 中 $\mathrm{Var}(X)$ 和 $\mathrm{Var}(Y)$.

38. 如果 $E[X]=1$ 及 $\mathrm{Var}(X)=5$, 计算

(a) $E[(2+X)^2]$. (b) $\mathrm{Var}(4+3X)$.

39. 从一个装有 3 个白球 3 个黑球的坛子里随机地取球, 然后放回, 再取, 一直进行下去. 问最先取出的 4 个球当中, 恰有 2 个是白球的概率?

40. 一次考试有 5 道题目, 同时每道题列出 3 个可能答案, 其中有一个答案是正确的. 某学生靠猜测能答对至少 4 道题的概率是多大?

41. 某人自称有超感官力. 作为对他的考验, 将一枚均匀的硬币抛 10 次, 让他事先预测抛得的结果, 10 次中他说对了 7 次. 如果他没有超感官力, 他将作出至少这样好的答案的概率是多少?

42. 飞机在飞行中引擎出故障的概率为 $1-p$, 且各引擎独立. 如果飞机要半数以上引擎工作正常才能安全飞行, 那么 p 为多大值时, 5 个引擎的飞机要优于 3 个引擎的飞机?

43. 某通讯系统传送数字 0 与 1, 但由于静电干扰, 传送的数字被错误地接受的概率为 0.2. 假定我们要传送一个由 0 和 1 组成的重要电报, 为减少出错的机会, 我们用 00000 代替 0, 用 11111 代替 1. 如果收讯者用过半译码法 (即收到一半以上为 0 则译码为 0, 一半以上为 1 则译码为 1), 那么被传送的每个信息译出后是错误的概率有多大? 此处作了何种独立性假设?

44. 某卫星系统由 n 个元件组成, 在某天内如果至少 k 个元件工作正常, 那么卫星工作正常. 某下雨天, 每个元件失效的概率都为 p_1, 且相互独立. 而天晴时每个元件失效的概率为 p_2, 也相互独立. 如果明天下雨的概率为 α, 那么明天卫星仍正常工作的概率是多大?

45. 某学生已准备好参加一次重要的口试, 且对口试那天他将处于怎样的精神状态很关心. 因为他估计, 如果他处于最佳状态, 则其主考人各自独立地给他及格的概率为 0.8, 若他处于很坏的状态, 上述概率降到 0.4. 若假定有过半数主考人给他及格, 他的口试才算合格, 且这个学生已感到他处于很坏的状态的可能性比处于最佳状态的可能性大一倍, 问他是邀请 5 位主考人还是 3 位主考人为好?

46. 假定某 12 个人组成的陪审团至少有 9 票有罪票才能判决一被告有罪, 并设陪审员对有罪人投无罪票的概率为 0.2, 而对无罪人投有罪票的概率为 0.1, 如果各陪审员的行为相互独立, 且 65% 的被告是事实上有罪的. 试求此陪审团作出正确判决的概率. 被告中被判为有罪的占多大百分比?

47. 某些军事法庭常任命 9 名法官, 原告与被告的律师都有权拒绝任何法官出庭, 被拒绝的法官即离庭且不再找人替换. 如果过半数法官投有罪票, 则被告被宣判有罪, 否则, 他就被判为无罪. 假设对一个事实上有罪的被告, 各法官 (独立地) 投有罪票的概率为 0.7, 而对事实上无罪的被告, 这个概率下降到 0.3.

(a) 当出庭法官数如下时, 一个有罪的被告被判为有罪的概率是多少?

(i) 9;　(ii) 8;　(iii) 7.

(b) 对一个无罪的被告重做 (a).

(c) 假定原告律师不行使拒绝法官出庭的权利, 并限定被告律师至多可拒绝两名法官, 如果被告律师有 60% 的把握判断他的当事人事实上是有罪的, 那么他拒绝几名法官为好?

48. 已知某公司生产的磁盘为残次品的概率为 0.01, 且相互独立. 该公司按包出售磁盘, 每包 10 张, 并且作出保证: 每包里最多有一个残次品磁盘, 否则予以退货处理. 如果某人买了 3 包, 问正好要退货 1 包的概率是多大?

49. 当掷硬币 1 时, 正面朝上的概率为 0.4; 当掷硬币 2 时, 正面朝上的概率为 0.7. 随机选一枚硬币掷 10 次:

(a) 10 次投掷中, 正好有 7 次正面朝上的概率是多大?

(b) 在第一次投掷为正面朝上的条件下, 在 10 次投掷中正好有 7 次正面朝上的条件概率是多大?

50. 假设掷一枚不均匀硬币, 正面朝上的概率为 p, 连续掷 10 次, 已知一共 6 次正面朝上, 求以下事件的条件概率:

(a) 前三次为 H,T,T(意味着第一次为正面朝上, 第二次为反面朝上, 第三次为反面朝上);

(b) 前三次为 T,H,T.

51. 某本杂志的一页上的印刷错误的个数的期望为 0.2, 那么下一页的印刷错误数为 (a) 0, (b) 2 的概率是多大? 解释理由.

52. 全世界每个月商业飞机发生坠毁事故的平均值为 3.5, 以下事件的概率是多大?

(a) 下个月至少有 2 起坠毁事故.　(b) 下个月至多 1 次坠毁事故.

试解释原因.

53. 去年纽约州大概举行 80 000 次婚礼. 估计如下概率值:

(a) 至少有一对夫妇他们都出生于 4 月 30 日,　(b) 至少有一对夫妇生日相同.

说明你做的假设.

54. 某高速公路上每周丢弃车辆的平均值为 2.2, 求以下事件概率的近似值:

(a) 下一周没有丢弃车辆;　(b) 下一周至少丢弃两辆车.

55. 某打字社雇了两名打字员. 第一个打字员打字时, 每篇文章出错的平均数目为 3, 而第二个打字员每篇文章出错的平均数目为 4.2. 如果你的文章等可能地分配给这两个人, 近似估计没有出错的概率.

56. 至少需要多少人, 这样其中至少有一人与你生日相同的概率超过 1/2?

57. 假设某高速公路上每天事故数是一参数为 $\lambda = 3$ 的泊松随机变量,

(a) 求今天至少发生 3 件事故的概率;

(b) 在今天至少发生了一件事故的假定条件下, 重做 (a).

58. 在下列情形下, 比较泊松近似和二项随机变量的概率:

(a) $P\{X=2\}$, 其中 $n=8, p=0.1$;　(b) $P\{X=9\}$, 其中 $n=10, p=0.95$;

(c) $P\{X=0\}$, 其中 $n=10, p=0.1$;　(d) $P\{X=4\}$, 其中 $n=9, p=0.2$.

59. 如果你买了 50 张彩票, 每张中彩的机会为 1/100, 问你将有以下中彩数的 (近似) 概率是多少?

(a) 至少 1 张;　(b) 正好 1 张;　(c) 至少 2 张.

60. 给定一年内, 人患感冒的次数为参数为 $\lambda=5$ 的泊松随机变量. 假设有种新药物 (基于大量维生素 C) 经过市场验证, 对 75% 的人有效, 并能将泊松分布的参数减少为 $\lambda=3$, 对另外 25% 的人不起作用. 如果某个人服用了该药, 一年内患了 2 次感冒, 那么该药对他 (她) 有效的可能性是多大?

61. 一手牌分到福尔豪斯的概率约为 0.0014, 如果你玩 1000 手牌, 求至少有 2 手福尔豪斯的近似概率.

62. 考虑 n 次相互独立的试验, 每次试验有 k 个可能结果, 其相应的概率为 $p_1, \cdots, p_k, \sum_{i=1}^k p_i = 1$. 若所有的 p_i 都比较小, 指出 n 个试验中, 没有两个试验结果相同的概率近似地等于 $\exp\{-n(n-1)\sum_{i=1}^n p_i^2/2\}$.

63. 进入某赌场的速度为每两分钟一人, 试问:

(a) 在 12:00 至 12:05 之间没有人进入赌场的概率是多大?

(b) 在这段时间内, 至少有 4 人进入赌场的概率是多大?

64. 某州每个月的自杀率为每 100 000 个居民中有一个自杀.

(a) 该州的一个拥有 400 000 居民的城市, 在给定月份内至少有 8 名自杀者的概率是多大?

(b) 一年内至少有两个月, 每月至少有 8 名自杀者的概率是多大?

(c) 记当前的月份为 1, 第一次超过 8 名自杀者的月份为 $i, i \geqslant 1$ 的概率是多大?
你作了哪些假设?

65. 某军营有 500 士兵, 每人得某种疾病的概率为 $1/10^3$, 且相互独立. 为了方便, 将 500 份血样混在一起进行检测.

(a) 求混合的血液呈阳性的近似概率 (这样至少有一名士兵患有此疾病).

现假定血液呈阳性.

(b) 这种情况下, 多于 1 人患此疾病的概率有多大?

再设其中一人琼斯知道自己患有该疾病.

(c) 琼斯认为多于 1 人患此病的概率有多大?

由于混合样本为阳性, 医生决定每一个人都要测试. 前面 $i-1$ 个人为阴性, 第 i 人为琼斯, 检查为阳性.

(d) 作为 i 的函数, i 后面有人患此病的概率是多大?

66. 由 n 对夫妇组成的 $2n$ 个人随机 (任何一种顺序都是等可能的) 坐在一张圆桌上. 记 C_i 为“第 i 对夫妻坐在一起”, $i=1, \cdots, n$.

(a) 求 $P(C_i)$.　(b) 对 $j \neq i$, 求 $P(C_j|C_i)$.

(c) 当 n 很大时, 近似计算没有一对夫妻坐在一起的概率.

67. 重复计算上面的问题, 如果要求男女间隔坐.

68. 为了对付 10 枚导弹的袭击, 发射了 500 枚拦截导弹. 拦截导弹随机地选择导弹作为目标, 而且是等可能的. 如果拦截导弹命中目标的概率是 0.1, 且相互独立, 利用泊松范例近似计算所有导弹被成功拦截的概率.

69. 掷一枚均匀硬币 10 次, 用以下方法计算存在一个长度为 4 的正面朝上序列的概率:

(a) 利用课文中导出的公式; (b) 利用课文中导出的递推公式;

(c) 与用泊松近似得到的结果进行对比.

70. 在时刻 0 时投掷一枚正面朝上概率为 p 的硬币, 设它落地后也正面朝上. 根据参数为 λ 的泊松过程选定时刻捡起硬币并抛掷 (两次抛掷之间硬币就放在地上). 那么在时刻 t 硬币为正面朝上的概率是多大?

提示: 到时刻 t 没有额外的投掷的条件下, 其条件概率是多大? 如果有额外投掷, 其条件概率又是多大?

71. 考虑轮盘赌中的一个轮盘, 有 38 个数字 ——从 1 到 36 还有 0 以及 00. 如果史密斯经常押注在 1 到 12 之间, 试求以下概率:

(a) 史密斯将要输掉开始的 5 局; (b) 他第一次赢钱将是第 4 次下注.

72. 两个球队进行一系列比赛, 先赢得 4 场比赛的球队获得最终胜利. 假设其中一球队的水平比另一球队稍强, 其赢得每一场胜利的概率为 0.6, 且各场的胜负相互独立. 求强队获得最终胜利时比赛的场数为 i 的概率, $i = 4, 5, 6, 7$. 并与在 3 局两胜的比赛中获胜的概率进行比较.

73. 考虑习题 72, 如果两队势均力敌, 每场比赛每队获胜的概率为 1/2, 计算比赛结束时比赛场数的期望值.

74. 某记者有一份要采访的人的名单, 假设该记者需要采访 5 人, 且任一个人 (独立地) 同意被采访的概率为 2/3. 那么名单上的人数为(a)5 人(b)8 人时, 他能达到采访人数要求的概率为多大?

对于 (b) 来说, 该记者采访到其名单上的(c)6 人(d)7 人时才刚好达到采访人数要求的概率为多大?

75. 连续掷一枚均匀硬币, 直到出现第 10 次正面朝上为止. 令 X 表示反面朝上出现的次数, 计算 X 的分布列.

76. 求解以下情形的巴拿赫火柴问题 (例 8e): 在开始时左边的火柴盒有 N_1 根火柴, 右边的火柴盒有 N_2 根火柴.

77. 在巴拿赫火柴问题中, 求以下事件概率: 第一个盒子取空时 (这个时刻与发现盒子为空盒的时刻是不一样的!), 另一个盒子还有 k 根火柴.

78. 坛子里有 4 个白球, 4 个黑球, 随机从中抽取 4 个球, 如果 2 黑 2 白, 那么停止抽球. 否则, 将这些球放回再次抽取 4 个, 直到抽取的 4 个球中正好有 2 个白球, 那么此时我们正好做了 n 次抽取的概率是多大?

79. 假设一批 100 个零件中有 6 个残次品, 其他 94 个是合格品. 如果随机从中抽取 10 个, 记 X 为其中“残次品数”, 求 (a) $P\{X = 0\}$, (b) $P\{X > 2\}$.

80. 流行于内华达州赌场的一种赌博叫“凯诺”, 其赌法如下: 赌场从 1 至 80 号中随机地取出 20 个号码, 赌徒再从这 80 个号码中任取 1 至 15 个号码. 如果赌徒取出的号码中有一定

比例个数的号码与赌场所取出的号码相同, 则他取胜, 赢得的钱数是赌徒取号码的个数与其中配上对的号码个数的函数. 例如, 赌徒只取一个号码时, 若此号码在赌场所取的 20 个号码中, 则他每下注 1 元赌注赢 2.2 元. (由于此时赌徒的获胜概率为 1/4, 所以公平的赢得钱数显然应是每 1 元赌注赢 3 元.) 当赌徒取出 2 个号码时, 若这 2 个号码都在赌场所取的 20 个号码之中, 则每下 1 元赌注赢得 12 元.

(a) 此时公平的赢钱数应是多少?

令 $P_{n,k}$ 表示赌徒取出 n 个号码之中恰有 k 个在赌场取出的 20 个号码之中的概率.

(b) 求出 $P_{n,k}$.

(c) 在凯诺赌中, 最典型的赌法是赌徒取 10 个号码, 这种情况下赌场付出的钱数如下表. 试计算期望赢得的钱数.

押注 10 个号码赢得钱数

猜对个数	每一元赌注赢得钱数
0~4	−1
5	1
6	17
7	179
8	1299
9	2599
10	24 999

81. 在例 8i, 具有 i 个残次品的包被拒绝的概率有多大? 计算 $i=1$ 及 $i=4$ 时的拒绝概率. 如果已知一个包被拒绝了, 那么这个包含有 4 个残次品的条件概率是多大?

82. 购买者按包购买某晶体管, 一包 20 个, 其购买方式如下: 随机检查该包里 4 个, 如果 4 个全是合格品, 那么接受这包, 否则就拒绝. 如果每包里的零件是否为残次品是相互独立的, 且概率为 0.1, 那么拒绝的包数的比例是多大?

83. 一个省里有 3 条高速公路, 每天在高速公路上发生的事故数是泊松随机变量, 其参数分别为 0.3, 0.5 和 0.7. 找出今天在高速公路上发生事故总数的期望值.

84. 有 10 个球要放到 5 个盒子中去. 每个球都是独立地被放入盒子中, 而放入第 i 个盒子的概率都是 $p_i, \sum_{i=1}^{5} p_i = 1$.

(a) 当 10 个球放完以后, 5 个盒子中可能有些盒子仍然是空的. 求空盒子数的期望值.

(b) 当 10 个球都放入盒子中去以后, 求只含 1 个球的盒子数的期望值.

85. 一共有 k 种优惠券, 每收集到一张优惠券时, 这张优惠券是第 i 种优惠券的概率为 p_i, $\sum_{i=1}^{k} p_i = 1$. 各次收集是相互独立的, 现在假定已经收集到 n 张优惠券, 问收集到优惠券的种类数的期望值是多少?

理论习题

1. 有 N 种不同的优惠券, 每次收集的种类都是相互独立的, 而且收集到第 i 种的概率为 $P_i, i=1,2,\cdots,N$. 令 T 表示某人为了每个种类至少收集到一张需要的总张数, 计算

$P\{T=n\}$.

提示：利用类似例 1e 的方法.

2. 如果 X 的分布函数为 F, 那么 e^X 的分布函数是什么?

3. 如果 X 的分布函数为 F, 那么随机变量 $\alpha X+\beta$ 的分布函数是什么? 其中 α,β 是常数, $\alpha\neq 0$.

4. 对取值为非负整数的随机变量 N, 证明

$$E[N]=\sum_{i=1}^{\infty}P\{N\geqslant i\}.$$

提示：$\sum_{i=1}^{\infty}P\{N\geqslant i\}=\sum_{i=1}^{\infty}\sum_{k=i}^{\infty}P\{N=k\}$. 然后交换求和次序.

5. 对一个取值非负整数的随机变量 N, 证明

$$\sum_{i=0}^{\infty}iP\{N>i\}=\frac{1}{2}(E[N^2]-E[N]).$$

提示：$\sum_{i=0}^{\infty}iP\{N>i\}=\sum_{i=0}^{\infty}i\sum_{k=i+1}^{\infty}P\{N=k\}$, 然后交换求和次序.

6. 设 X 满足

$$P\{X=1\}=p=1-P\{X=-1\}$$

求 $c,c\neq 1$, 使得 $E[c^X]=1$.

7. 设 X 为一随机变量, 期望值为 μ, 方差为 σ^2. 求以下随机变量的期望和方差:

$$Y=\frac{X-\mu}{\sigma}$$

8. 设 X 满足

$$P(X=a)=p=1-P(X=b)$$

求 $\mathrm{Var}(X)$.

9. 利用二项概率

$$P\{X=i\}=\binom{n}{i}p^i(1-p)^{n-i},\quad i=0,\cdots,n$$

导出二项展开公式

$$(x+y)^n=\sum_{i=0}^{n}\binom{n}{i}x^iy^{n-i}$$

其中 x,y 为非负数. 提示：令 $p=\dfrac{x}{x+y}$.

10. 设 X 为服从参数为 (n, p) 的二项分布的随机变量, 证明:

$$E\left[\frac{1}{X+1}\right]=\frac{1-(1-p)^{n+1}}{(n+1)p}$$

11. 考虑 n 重独立重复试验, 每次成功的概率为 p. 证明 k 次成功, $n-k$ 次失败的共 $n!/[k!(n-k)!]$ 种可能排列方式都是等可能的.

12. n 个元件排成一排, 每个元件工作正常的概率为 p, 且各元件相互独立. 问没有两个相邻的元件都工作不正常的概率是多大?

提示：以工作不正常的元件的个数作为条件, 利用第 1 章例 4c 的结果.

13. X 为服从参数为 (n,p) 的二项分布的随机变量, p 取何值时能使 $P\{X=k\}$ 取最大值? $k=0,1,\cdots,n$. 这是关于一个统计方法的例子, 是当观察到 $X=k$ 时, 估计未知参数 p 的一种统计方法. 如果我们假定 n 已知, 通过选择使得 $P\{X=k\}$ 最大来估计 p, 这就是所谓的极大似然估计方法.

14. 一个家庭有 n 个孩子的概率为 $\alpha p^n, n\geqslant 1$, 其中 $\alpha\leqslant(1-p)/p$.

(a) 没有孩子的家庭的比例是多大?

(b) 如果孩子是男孩或女孩的概率相等, 且各孩子的性别是相互独立的. 那么有 k 个男孩 (可能还有若干女孩) 的家庭比例是多大?

15. 考虑 n 次独立重复投掷一枚硬币, 每次正面朝上的概率为 p. 证明有偶数个正面朝上的概率为 $[1+(q-p)^n]/2$, 其中 $q=1-p$. 此题, 要先证明再利用如下等式:

$$\sum_{i=0}^{[n/2]}\binom{n}{2i}p^{2i}q^{n-2i}=\frac{1}{2}[(p+q)^n+(q-p)^n]$$

其中 $[n/2]$ 是不超过 $n/2$ 的最大整数. 将本题与第 3 章的理论习题 16 进行比较.

16. 令 X 为服从参数为 λ 的泊松分布的随机变量, 证明 $P\{X=i\}$ 随着 i 的增加, 先是单调递增, 然后再单调递减, 当 i 取值为不超过 λ 的最大整数时, 它取得最大值.

提示: 考虑 $P\{X=i\}/P\{X=i-1\}$.

17. 令 X 为服从参数为 λ 的泊松分布的随机变量.

(a) 证明:

$$P\{X\text{为偶数}\}=\frac{1}{2}\left[1+\mathrm{e}^{-2\lambda}\right]$$

利用理论习题 15 的结果以及泊松分布和二项分布之间的关系证明.

(b) 直接利用 $\mathrm{e}^{-\lambda}+\mathrm{e}^{\lambda}$ 的展开式来证明上述等式.

18. 令 X 为服从参数为 λ 的泊松分布的随机变量, λ 取何值时 $P\{X=k\}$ 取最大值 $(k\geqslant 0)$.

19. 如果 X 为服从参数为 λ 的泊松分布的随机变量, 证明

$$E[X^n]=\lambda E[(X+1)^{n-1}]$$

利用该结果计算 $E[X^3]$.

20. 考虑 n 枚硬币, 每枚硬币正面朝上的概率都是 p, 且相互独立. 假设 n 较大, 而 p 较小, 令 $\lambda=np$. 假设将这 n 枚硬币一起抛掷, 一直到至少有一枚硬币正面朝上为止, 否则继续掷这 n 枚硬币. 也即, 停下来的时候至少有一枚硬币正面朝上. 令 X 表示停下来时正面朝上的出现的总枚数. 下列关于 $P\{X=1\}$ 的近似计算的理由哪个是正确的? 所有情形中, Y 都是一个服从参数为 λ 的泊松随机变量.

(a) 因为掷 n 枚硬币时, 正面朝上出现的次数近似于服从参数为 λ 的泊松分布, 因此

$$P\{X=1\}\approx P\{Y=1\}=\lambda\mathrm{e}^{-\lambda}$$

(b) 因为掷 n 枚硬币时, 正面朝上出现的次数近似于服从参数为 λ 的泊松分布, 而且此数为正时才会停止, 因此

$$P\{X=1\}\approx P\{Y=1|Y>0\}=\frac{\lambda\mathrm{e}^{-\lambda}}{1-\mathrm{e}^{-\lambda}}$$

(c) 因为至少出现一枚正面朝上, 当其他 $n-1$ 枚都不是正面朝上时 $X=1$. 又因 $n-1$ 枚硬币出现的正面朝上的枚数近似服从参数为 $(n-1)p \approx \lambda$ 的泊松分布, 因此

$$P\{X=1\} \approx P\{Y=0\} = \mathrm{e}^{-\lambda}$$

21. 随机抽取了 n 人, 令 E_{ij} 表示事件“第 i 人和第 j 人生日相同”, 假定一个人的生日在 365 天任一天的可能性是一样的. 计算:

(a) $P(E_{3,4}|E_{1,2})$; (b) $P(E_{1,3}|E_{1,2})$; (c) $P(E_{2,3}|E_{1,2} \cap E_{1,3})$.

关于 $\binom{n}{2}$ 个事件 E_{ij} 的独立性, 你从上能得出什么结论?

22. 坛子里有 $2n$ 个球, 其中两个标号 1, 两个标号 2, $\cdots$, 两个标号 n. 每次无放回地随机从中取出 2 个, 令 T 表示取出来的两个球第一次同号时的取球次数, (如果根本没有出现过两球同号, 那么 $T=\infty$), 对 $0<\alpha<1$, 我们要证明

$$\lim_{n\to\infty} P\{T>\alpha n\} = \mathrm{e}^{-\alpha/2}$$

为了证明这点, 令 M_k 表示在前 k 次取球后取出来的号码相同的球对数, $k=1,\cdots,n$.

(a) 说明当 n 很大时, M_k 可认为是k 次 (近似) 独立重复试验中试验成功的次数.

(b) 当 n 很大时, 求 $P\{M_k=0\}$ 的近似值.

(c) 利用随机变量 M_k 的值表示事件 $\{T>\alpha n\}$.

(d) 给出 $P\{T>\alpha n\}$ 的公式, 求极限概率.

23. 考虑有 n 个人 (n 至少取值在 $80\sim 90$ 之间), 在近似计算没有 3 个人同一天生日的概率时, 一个比在文中得到的要好的泊松近似如下: 令 E_i 表示事件“至少有 3 人生日在第 i 天”, $i=1,\cdots,365$.

(a) 计算 $P(E_i)$; (b) 给出没有 3 人同一天生日的概率的近似值;

(c) 当 $n=88$(可以证明 $n=88$ 为使得概率超过 1/2 的最小 n 的值) 时, 计算上述概率.

24. 连续掷 n 枚硬币, 每次正面朝上的概率为 p. 考虑事件“试验序列中出现连续 k 次正面朝上”的概率 P_n. 以下是计算这个概率的另一个方法.

(a) 证明, 对于 $k<n$, 在试验序列中出现连续的 k 次正面朝上的充要条件是:

(i) 在前面 $n-1$ 次掷硬币试验结果序列中, 已经出现连续的 k 次正面向上的子序列, 或者

(ii) 在前面 $n-k-1$ 次掷硬币的试验结果序列中, 没有出现连续 k 次正面朝上的子序列, 但第 $n-k$ 次掷硬币时正面朝下, 第 $n-k+1,\cdots,n$ 次都是正面向上.

(b) 利用 (a) 建立 P_n 与 P_{n-1} 的递推关系. 从 $P_k=p^k$ 开始, 利用该递推公式可得 P_{k+1}, 然后 P_{k+2}, 等等, 直到 P_n.

25. 假设在某个时间内事件发生的次数为参数为 λ 的泊松随机变量. 如果每个事件被统计到的概率为 p, 且各个事件是否被统计是相互独立的, 指出被统计到的事件的数目为参数为 λp 的泊松随机变量, 并给出一个直观解释. 作为上述理论的应用, 假设某个地区的铀的矿点数为参数为 $\lambda=10$ 的泊松随机变量. 如果, 给定一个时间段内, 每个矿点被发现的概率为 1/50, 且相互独立. 求以下事件的概率:

(a) 这段时间内正好发现 1 个矿点; (b) 这段时间内至少发现 1 个矿点;

(c) 这段时间内最多发现 1 个矿点.

26. 证明:

$$\sum_{i=0}^{n} \mathrm{e}^{-\lambda}\frac{\lambda^i}{i!} = \frac{1}{n!}\int_{\lambda}^{\infty} \mathrm{e}^{-x}x^n \mathrm{d}x$$

提示: 利用分部积分.

27. 如果 X 为一几何随机变量, 给出分析的证明:

$$P\{X = n + k|X > n\} = P\{X = k\}$$

利用几何随机变量的定义直观地说明上式成立的原因.

28. 令 X 为参数为 (r, p) 的负二项随机变量, 令 Y 为参数为 (n, p) 的二项随机变量, 指出

$$P\{X > n\} = P\{Y < r\}$$

提示: 可以有两种方法完成证明. 一种用分析的方法, 上式等价于下列恒等式

$$\sum_{i=n+1}^{\infty}\binom{i-1}{r-1}p^r(1-p)^{i-r} = \sum_{i=0}^{r-1}\binom{n}{i}p^i(1-p)^{n-i}$$

另一种是利用随机变量的概率解释. 即考虑进行一系列成功率均为 p 的独立试验, 用试验结果表示事件 $\{X > n\}$ 和 $\{Y < r\}$.

29. 设 X 为一超几何随机变量, 计算 $P\{X = k+1\}/P\{X = k\}$.

30. 坛子里装有标有号码 1 到 N 的球. 假设随机无放回地抽取 n 个球, $n \leqslant N$. 令 Y 表示抽取的球中最大号码.

(a) 求 Y 的分布列.

(b) 导出 $E[Y]$ 的表达式, 然后利用费马组合恒等式 (参见第 1 章理论习题 11) 予以简化.

31. 坛子里有 $m+n$ 个芯片, 分别标有号码 $1, 2, \cdots, n+m$. 从中取出 n 个. 如果令 X 表示抽取的芯片中, 其号码大于留在坛子中芯片的最大号码的数量, 求 X 的分布列.

32. 坛子里有 n 个芯片. 某男孩连续有放回地随机从中抽取, 一直进行下去, 直到他取出一个前面曾经取出过的芯片为止. 令 X 表示此时的抽取次数, 计算其分布列.

33. 利用 (7.5) 式推导 (7.6) 式.

34. 从一 n 个元素的集合里, 随机选择一个非空子集. 假定所有的非空子集被选中的可能性是一样的. 令 X 表示选中的子集的元素的个数. 利用第 1 章理论习题 12 给出的恒等式, 证明

$$E[X] = \frac{n}{2-\left(\frac{1}{2}\right)^{n-1}} \qquad \mathrm{Var}(X) = \frac{n\cdot 2^{2n-2} - n(n+1)2^{n-2}}{(2^n-1)^2}$$

证明, 当 n 很大时,

$$\mathrm{Var}(X) \sim \frac{n}{4}$$

上式的意义是当 $n \to \infty$ 时 $\mathrm{Var}(x)/(n/4) \to 1$. 比较这个结果与 $\mathrm{Var}(Y)$, 其中 Y 是离散集 $\{1, \cdots, n\}$ 上的均匀随机变量, 即 $P\{Y = i\} = 1/n, i = 1, \cdots, n$.

35. 坛子里最初有一个红球, 一个蓝球. 每次随机取一个球, 将该球放回坛子的同时还放进另一个同颜色的球. 令 X 表示第一次抽取到蓝色的球时所抽取的次数. 比如, 如果第一次拿出了红球, 第二次拿出的是蓝球, 那么 X=2.

(a) 求 $P\{X > i\}, i \geqslant 1$.

(b) 证明, 最终能取出蓝球的概率为 1. (也即, 证明 $P\{X < \infty\} = 1$.) (c) 计算 $E[X]$.

36. 设 X 的取值范围为 $\{x_i\}$, Y 的取值范围为 $\{y_j\}$, $Z = X + Y$ 的取值范围为 $\{z_k\}$. 记 A_k 为满足 $x_i + y_j = z_k$ 的指标 (i, j) 的集合, 即 $A_k = \{(i, j) : x_i + y_j = z_k\}$.

(a) 证明

$$P\{X + Y = z_k\} = \sum_{(i,j)\in A_k} P\{X = x_i, Y = y_j\}$$

(b) 证明

$$E[X + Y] = \sum_k \sum_{(i,j)\in A_k} (x_i + y_j)P\{X = x_i, Y = y_j\}$$

(c) 利用 (b) 中的公式证明

$$E[X + Y] = \sum_i \sum_j (x_i + y_j)P\{X = x_i, Y = y_j\}$$

(d) 证明

$$P(X = x_i) = \sum_j P\{X = x_i, Y = y_j\}$$

$$P(Y = y_j) = \sum_i P\{X = x_i, Y = y_j\}$$

(e) 证明

$$E[X + Y] = E[X] + E[Y]$$

自 检 习 题

1. 考虑某个棒球运动员在他接下来的 3 次击打中所获得的分数, 令 X 表示此随机变量, X 的可能取值为 0, 1, 2, 3, 如果 $P\{X = 1\} = 0.3, P\{X = 2\} = 0.2$, 且 $P\{X = 0\} = 3P\{X = 3\}$, 求 $E[X]$.

2. 假设 X 的取值为 $0, 1, 2$ 之一, 如果对某个常数 c, 有 $P\{X = i\} = cP\{X = i - 1\}, i = 1, 2$, 求 $E[X]$.

3. 投掷一枚硬币, 每次正面朝上的概率为 p, 连续投掷直到正面或反面朝上出现两次, 求投掷总次数的期望.

4. 某个社区由 m 个家庭组成, 其中有 i 个孩子的家庭有 n_i 个: $\sum_{i=1}^{r} n_i = m$. 随机挑选一个家庭, 令 X 表示该家庭的孩子数. 再随机从 $\sum_{i=1}^{r} in_i$ 个孩子中随机挑选一个孩子, 令 Y 表示该孩子所在的家庭里的孩子数, 证明: $E[Y] \geqslant E[X]$.

5. 假设 $P\{X = 0\} = 1 - P\{X = 1\}$. 如果 $E[X] = 3\text{Var}(X)$, 求 $P\{X = 0\}$.

6. 箱子里有两枚硬币. 抛掷其中一枚硬币, 正面朝上的概率为 0.6. 抛掷另一枚硬币, 正面朝上的概率为 0.3. 随机地从箱子中拿出一枚硬币并抛掷, 在不知道拿的是哪一枚硬币的情况下, 你可以押至多 10 元的注: 如果是正面朝上, 你将赢得你押的数量; 如果是反面朝上, 你就输掉押注. 然而, 假设有一位内幕人, 他想卖给你信息, 告诉你到底是选中了哪枚硬币, 这个信息卖 C 元. 如果你买了这个信息, 你的期望回报是多少? 注意到如果你买了它, 然后押注 x 元, 那么最后你可能赢得 $x-C$ 元, 或 $-x-C$ 元 (即你输掉 $x+C$ 元), 那么应该为这条信息支付多少呢?

7. 一个慈善家在一张红纸上写了一个数字 x, 给一个公证人看后, 把纸片翻过来放在桌子上. 公证人然后掷一枚均匀的硬币, 如果是正面朝上, 他在一张蓝纸上写下 $2x$, 如果是反面朝上, 写下 $x/2$, 然后将纸也翻过来放在桌子上. 在不知道 x 的值, 也不知道掷硬币的结果的情况下, 你可以选择翻开红纸还是翻开蓝纸. 看了翻开纸上的数值后, 你可以决定是要这张翻开的纸上写的数值作为奖励, 还是要另外一张纸上写的数值 (此时不知道大小) 作为奖励. 比如, 如果你决定翻开蓝纸, 看了数值为 100, 这时你可以选择 100 作为你的奖励, 也可以选择红纸上的数值 (50 或 200) 作为奖励. 设想你希望奖励的期望值最大.

(a) 指出没有理由先翻开红纸, 因为如果这样做的话, 那么不管红纸上的值是多大, 最好还应该再翻开蓝纸;

(b) 令 y 为一个固定的非负数, 考虑如下策略: 翻开蓝纸, 如果上面的数值至少是 y, 那么就接受这个奖励; 如果比 y 小, 那么再翻开红纸. 如果慈善家写的是 x, 令 $R_y(x)$ 表示你执行该策略最后获得的奖励数, 计算 $E[R_y(x)]$. 注意到 $E[R_0(x)]$ 是慈善家写的是 x, 而你执行“始终选择蓝纸”这一策略所获得的期望奖励.

8. 令 $B(n,p)$ 表示服从参数为 (n, p) 的二项分布的随机变量, 说明:

$$P\{B(n,p) \leqslant i\} = 1 - P\{B(n, 1-p) \leqslant n-i-1\}$$

提示: 将“成功的次数小于或等于 i”写成与之等价的关于失败次数的陈述.

9. 如果 X 是一个服从二项分布的随机变量, 期望为 6, 方差为 2.4, 求 $P\{X=5\}$.

10. 一个坛子里有 n 个球, 标号从 1 到 n, 如果依次随机地有放回地取出 m 个球, 计算 $P\{X=k\}, k=1,\cdots,m$, 其中 X 表示抽取的 m 个球里最大的号码.

提示: 先计算 $P\{X \leqslant k\}$.

11. A 队和 B 队进行一系列比赛, 先胜三局者为比赛的获胜者. 假设每局 A 队获胜的概率为 p, 且各局相互独立. 求下列条件概率:

(a) 已知 A 队赢了第一局, 求它最终获胜的条件概率;

(b) 已知 A 最终获胜, 求它赢了第一局的条件概率.

12. 某个地方足球队将要参加 5 场比赛, 如果它在本周末的比赛中取得了胜利, 它将升入一个更高的级别中进行后四场比赛, 如果它本周末的比赛输掉了, 它就要降到一个更低的级别中进行后四场比赛. 在高级别比赛中, 该队每场比赛获胜的概率为 0.4, 且各场相互独立. 在低级别比赛中, 该队每场比赛获胜的概率为 0.7, 且各场相互独立. 如果该队赢得本周末比赛的概率为 0.5, 那么它在后来的 4 场比赛中至少获胜 3 场的概率是多大?

13. 一个 7 人陪审团里的每个成员作出正确决定的概率为 0.7, 且相互独立. 如果最后的决定采取少数服从多数的原则, 那么陪审团作出正确决定的概率是多大? 假设已知陪审团的 4 人意见相同, 那么陪审团作出正确决定的概率是多大?

14. 某个地区平均来说, 每年会遭遇 5.2 次飓风袭击, 那么今年遭遇飓风袭击的次数为 3 次或者更少的概率是多大?

15. 某个种类的昆虫产在一片树叶上的虫卵的数目服从参数为 λ 的泊松分布. 然而, 这样的随机变量的取值只有当它为正整数时才能知道, 因为如果是 0 的话, 我们不会记录这个数, 因为我们不知道这片树叶上是否有昆虫, 令 Y 表示观测到的虫卵数, 这样

$$P\{Y=i\}=P\{X=i|X>0\}$$

其中 X 是一个服从参数为 λ 的泊松分布的随机变量, 求 $E[Y]$.

16. 有 n 个男孩和 n 个女孩, 每人都随机且独立地选择一名异性. 如果正好有一名男孩和女孩互相选中了, 那么他们将被配成一对. 给女孩编上号码, 令 G_i 表示事件"号码为 i 的女孩被配成了一对", 令 $P_0=1-P(\bigcup_{i=1}^{n}G_i)$ 表示没有任何一对配成的概率.

(a) $P(G_i)$ 是多大? (b) $P(G_i|G_j)$ 是多大? (c) 当 n 足够大时, 求 P_0 的近似值.

(d) 当 n 足够大时, 求 P_k 的近似值, 它是正好配成了 k 对的概率.

(e) 利用关于事件和的概率的容斥恒等式来计算 P_0.

17. 有 n 对夫妇组成的 $2n$ 个人, 随机地被分成 n 组, 每组两人. 给妇女编上号码, 令 W_i 表示事件"第 i 个妇女正好与她丈夫分在一组".

(a) 求 $P(W_i)$.

(b) 对 $i\neq j$, 求 $P(W_i|W_j)$.

(c) 当 n 足够大时, 求没有妇女与她丈夫分在一组的概率的近似值.

(d) 如果在分组的时候规定必须是一个男人一个女人成为一个组, 那么上述问题有怎样的答案?

18. 某个赌场的顾客在玩轮盘赌时, 每次押 5 元在"红", 直到她一共赢了 4 次.

(a) 她一共押了 9 次的概率是多大? (b) 她停下来时赢得的期望是多大?

注释: 每次押注, 她将以 18/38 的概率赢得 5 元, 以 20/38 的概率输掉 5 元.

19. 有三个朋友去喝咖啡, 他们决定用掷硬币的方式来确定谁买单: 每人掷一枚硬币, 如果有人的结果与其他两人不一样, 那么由他买单. 如果三枚硬币的结果是一样的, 那么就重掷一轮. 一直这样下去, 直到确定了由谁来买单. 求以下事件概率:

(a) 正好进行了三轮就确定了由谁来买单;

(b) 进行了 4 轮以上才确定了由谁来买单.

20. 如果 X 是服从参数为 p 的几何分布的随机变量, 证明 $E[1/X]=\dfrac{-p\ln(p)}{1-p}$.

提示: 需要计算形如 $\sum_{i=1}^{\infty}a^i/i$ 的表达式的值. 要做到这一点, 可利用 $a^i/i=\int_0^a x^{i-1}\mathrm{d}x$, 然后交换积分和求和的顺序.

21. 假设

$$P\{X=a\}=p,\qquad P\{X=b\}=1-p$$

(a) 证明 $(X-b)/(a-b)$ 服从伯努利分布; (b) 计算 $\mathrm{Var}(X)$.

22. 你每局比赛获胜的概率为 p, 你计划玩 5 局, 但是如果你赢了第 5 局, 你就会继续玩, 直到你输掉一局.

(a) 求一共玩的局数的期望. (b) 求一共输掉的局数的期望.

23. 坛子里开始有 N 个白球 M 个黑球, 每次无放回地随机取出一个, 求在取出 m 个黑球之前已经取出 n 个白球的概率, $n \leqslant N, m \leqslant M$.

24. 将 10 个球放到 5 个盒子中去. 每次放球时与其他球的放置是相互独立的. 每个球放到盒子 i 的概率为 $p_i, \sum_{i=1}^{n} p_i = 1$. 记 X_i 为第 i 个盒子中所放的球数.

(a) X_i 的分布是什么? (尽可能具体)

(b) 对于 $i \neq j, X_i + X_j$ 是什么类型的随机变量.

(c) 求 $P\{X_1 + X_2 + X_3 = 7\}$.

25. 对于配对问题 (见第 2 章例 5m),

(a) 求所配的对数的期望值.

(b) 求所配的对数的方差.

26. 设 X 为几何随机变量, 其参数为 p, 并设 α 是 X 为偶数的概率.

(a) 利用 $\alpha = \sum_{i=1}^{\infty} P\{X = 2i\}$ 求出 α 的值.

(b) 利用条件 $X = 1$ 或 $X > 1$, 求出 α 的值.

第 5 章　连续型随机变量

5.1　简　　介

在第 4 章我们讨论了离散型随机变量, 这类随机变量的可能取值的个数或者是有限的, 或者是可数无限的. 然而, 还存在一类随机变量, 它们的可能取值是无限不可数的. 比如以下的两个例子: 火车到达某个车站的时间以及某个晶体管的寿命. 我们称 X 为一个连续型[①] (continuous) 随机变量, 如果存在一个定义在实数轴上的非负函数 f, 使得对于任一个实数集 B[②], 下式成立

$$P\{X \in B\} = \int_B f(x)\,\mathrm{d}x \tag{1.1}$$

函数 f 称为随机变量 X 的概率密度函数 (probability density function), 或者密度函数 (参见图 5.1).

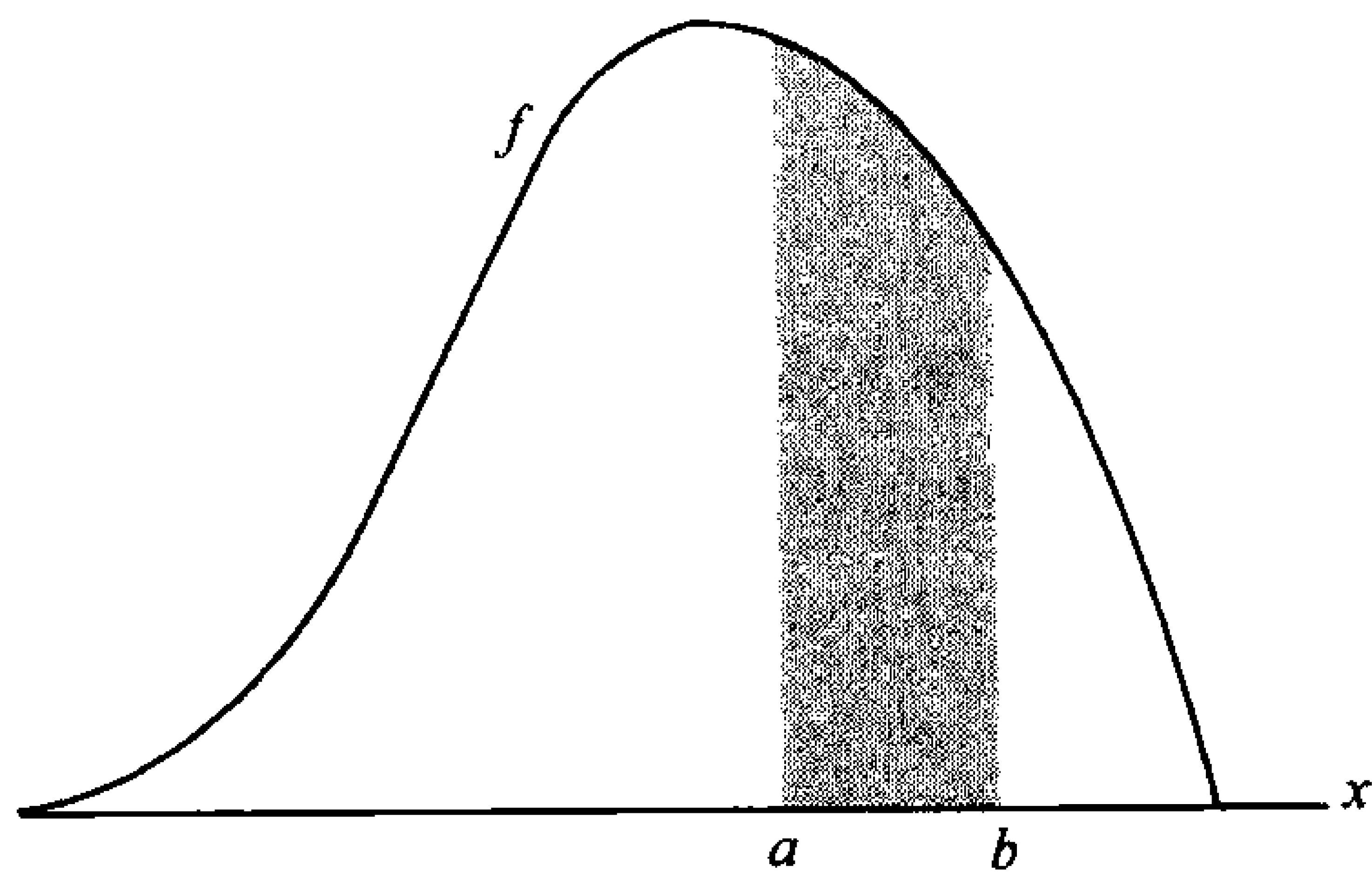


图 5.1　概率密度函数: $P(a \leqslant X \leqslant b)$ 等于阴影部分面积

换句话说, (1.1) 式说明了 X 属于 B 的概率可以通过对概率密度函数在集合 B 上积分得到. 既然 X 必取某个值, 因此 f 一定满足

$$1 = P\{X \in (-\infty, \infty)\} = \int_{-\infty}^{\infty} f(x)\mathrm{d}x$$

所有关于 X 的概率都可以通过 f 进行计算得到. 例如, 令 $B = [a, b]$, 通过 (1.1) 式可以得到

$$P\{a \leqslant X \leqslant b\} = \int_a^b f(x)\mathrm{d}x \tag{1.2}$$

在上式中令 $a = b$, 可以得到

$$P\{X = a\} = \int_a^a f(x)\mathrm{d}x = 0$$

也就是说, 对于一个连续型随机变量, 它取任何固定值的概率都等于 0. 因此, 对于

---

① 有时也称为绝对连续型(absolutely continuous).

② 事实上, (1.1) 式仅仅对可测集 B 成立. 当然, 很幸运, 现实中的集合都是可测集.

一个连续型随机变量, 有

$$P\{X < a\} = P\{X \leqslant a\} = F(a) = \int_{-\infty}^{a} f(x)\mathrm{d}x$$

例 1a 假设 X 是一个连续型随机变量, 其密度函数为

$$f(x) = \begin{cases} C(4x - 2x^2) & 0 < x < 2 \\ 0 & \text{其他} \end{cases}$$

(a) C 的值是多少? (b) 求 $P\{X > 1\}$.

解: (a) 既然 f 是一个概率密度函数, 那么一定有 $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$, 这意味着

$$C\int_0^2 (4x - 2x^2)\mathrm{d}x = 1 \quad \text{或者} \quad C\left[2x^2 - \frac{2x^3}{3}\right]\Bigg|_{x=0}^{x=2} = 1$$

这样 $C = 3/8$, 因此 (b) $P\{X > 1\} = \int_1^{\infty} f(x)\mathrm{d}x = \frac{3}{8}\int_1^2 (4x - 2x^2)\mathrm{d}x = \frac{1}{2}$ ■

例 1b 某台计算机在死机前连续运行的时间 (单位: 小时) 是一个连续型随机变量, 其密度函数为

$$f(x) = \begin{cases} \lambda \mathrm{e}^{-x/100} & x \geqslant 0 \\ 0 & x < 0 \end{cases}$$

以下事件的概率是多少?

(a) 该计算机在死机前运行的时间在 50 个小时到 150 小时之间;

(b) 运行时间不超过 100 小时.

解: (a) 因为

$$1 = \int_{-\infty}^{\infty} f(x)\mathrm{d}x = \lambda \int_0^{\infty} \mathrm{e}^{-x/100}\mathrm{d}x$$

这样可得

$$1 = -\lambda(100)\mathrm{e}^{-x/100}\Big|_0^{\infty} = 100\lambda \qquad \text{或} \qquad \lambda = \frac{1}{100}$$

因此, 电脑在死机前运行了 50 到 150 小时的概率为

$$\begin{aligned} P\{50 < X < 150\} &= \int_{50}^{150} \frac{1}{100}\mathrm{e}^{-x/100}\mathrm{d}x = -\mathrm{e}^{-x/100}\Big|_{50}^{150} \\ &= \mathrm{e}^{-1/2} - \mathrm{e}^{-3/2} \approx 0.384 \end{aligned}$$

(b) 类似地:

$$P\{X < 100\} = \int_0^{100} \frac{1}{100}\mathrm{e}^{-x/100}\mathrm{d}x = -\mathrm{e}^{-x/100}\Big|_0^{100} = 1 - \mathrm{e}^{-1} \approx 0.633$$

也就是说, 电脑在连续使用 100 小时以前, 大约 63.3% 的可能会死机. ■

例 1c　收音机的某种电子管的寿命是一随机变量, 其概率密度函数为

$$f(x)=\begin{cases}0 & x\leqslant 100\\ \dfrac{100}{x^2} & x>100\end{cases}$$

假定收音机里有 5 个电子管, 并且这些电子管的寿命是相互独立的, 在 150 小时内, 这台收音机的 5 个这样的电子管里正好有 2 个需要更换的概率是多大?

解: 令 E_i 表示"在给定时间内第 i 个电子管需要更换" $(i=1,2,\cdots,5)$,

$$P(E_i)=\int_0^{150}f(x)\mathrm{d}x=100\int_{100}^{150}x^{-2}\mathrm{d}x=\frac{1}{3}$$

利用事件 E_i 之间的独立性, 可得所求概率为

$$\binom{5}{2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^3=\frac{80}{243}$$

■

分布函数 F 与密度函数 f 之间的关系可以表示为

$$F(a)=P\{X\in(-\infty,a]\}=\int_{-\infty}^{a}f(x)\mathrm{d}x$$

对上式两边求导, 得到

$$\frac{\mathrm{d}}{\mathrm{d}a}F(a)=f(a)$$

也即, 密度函数是分布函数的导数. 从 (1.2) 式还可以得到一个关于概率密度函数更直观的解释:

$$P\left\{a-\frac{\varepsilon}{2}\leqslant X\leqslant a+\frac{\varepsilon}{2}\right\}=\int_{a-\varepsilon/2}^{a+\varepsilon/2}f(x)\mathrm{d}x\approx\varepsilon f(a)$$

其中 ε 是一个小数, 且 $f(\cdot)$ 在 $x=a$ 处连续. 换句话说, X 取值于以点 a 为中心, 长度为 ε 的小区间内的概率近似等于 $\varepsilon f(a)$. 通过这一点, 我们可以看出, $f(a)$ 是随机变量取值于点 a 附近的可能性的一个度量.

例 1d　设 X 是一个连续型随机变量, 其分布函数为 F_X, 密度函数为 f_X, 求 $Y=2X$ 的密度函数.

解: 用两种方法求解 f_Y. 第一种是直接求分布函数, 然后求导,

$$F_Y(a)=P\{Y\leqslant a\}=P\{2X\leqslant a\}=P\{X\leqslant a/2\}=F_X(a/2)$$

求导得到:

$$f_Y(a)=\frac{1}{2}f_X(a/2)$$

另一方法是, 注意到

$$\begin{aligned}\varepsilon f_Y(a)&\approx P\left\{a-\frac{\varepsilon}{2}\leqslant Y\leqslant a+\frac{\varepsilon}{2}\right\}=P\left\{a-\frac{\varepsilon}{2}\leqslant 2X\leqslant a+\frac{\varepsilon}{2}\right\}\\&=P\left\{\frac{a}{2}-\frac{\varepsilon}{4}\leqslant X\leqslant\frac{a}{2}+\frac{\varepsilon}{4}\right\}\approx\frac{\varepsilon}{2}f_X(a/2)\end{aligned}$$

两边除以 ε 就可以得到与前面相同的结果. ■

5.2 连续型随机变量的期望和方差

在第 4 章, 我们定义了离散型随机变量的期望值如下

$$E[X]=\sum_x xP\{X=x\}$$

如果 X 是一个连续型随机变量, 其密度函数为 $f(x)$, 那么由

$$f(x)\mathrm{d}x\approx P\{x\leqslant X\leqslant x+\mathrm{d}x\}\qquad 对于很小的\ \mathrm{d}x$$

很容易看出, 可用类似的方法定义连续型随机变量的期望值为

$$E[X]=\int_{-\infty}^{\infty}xf(x)\mathrm{d}x$$

例 2a 随机变量 X 的密度函数为

$$f(x)=\begin{cases}2x & 如果\ 0\leqslant x\leqslant 1\\ 0 & 其他\end{cases}$$

求 $E[X]$.

解:

$$E[X]=\int xf(x)\,\mathrm{d}x=\int_0^1 2x^2\,\mathrm{d}x=\frac{2}{3}$$ ■

例 2b 随机变量 X 的密度函数为

$$f(x)=\begin{cases}1 & 如果\ 0\leqslant x\leqslant 1\\ 0 & 其他\end{cases}$$

求 $E[\mathrm{e}^X]$.

解: 令 $Y=\mathrm{e}^X$. 我们从计算 F_Y, 也即 Y 的分布函数开始. 对于 $1\leqslant x\leqslant \mathrm{e}$, 有

$$F_Y(x)=P\{Y\leqslant x\}=P\{\mathrm{e}^X\leqslant x\}=P\{X\leqslant \ln x\}=\int_0^{\ln x}f(y)\,\mathrm{d}y=\ln x$$

对 $F_Y(x)$ 求导, 我们可以推导出 Y 的概率密度函数如下:

$$f_Y(x)=\frac{1}{x}\qquad 1\leqslant x\leqslant \mathrm{e}$$

因此,

$$E[\mathrm{e}^X]=E[Y]=\int_{-\infty}^{\infty}xf_Y(x)\mathrm{d}x=\int_1^{\mathrm{e}}\mathrm{d}x=\mathrm{e}-1$$ ■

尽管上例中的方法是计算随机变量 X 的函数的期望值的常用的方法, 但正如离散型情形一样, 还存在另一种方法, 下面的命题就类似于第 4 章的命题 4.1.

命题 2.1 设 X 是一个连续型随机变量, 其概率密度函数为 $f(x)$, 那么对于任一实值函数 g, 有

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)\,\mathrm{d}x$$

在例 2b 中利用命题 2.1 可得

$$E[\mathrm{e}^X] = \int_0^1 \mathrm{e}^x \mathrm{d}x = \mathrm{e} - 1 \qquad 因为 f(x) = 1, 0 < x < 1$$

这个结果与例 2b 中的结果是一致的.

命题 2.1 的证明比离散情形下更复杂, 我们仅在随机变量 $g(X)$ 非负的条件下证明本命题. ($g(X)$ 为一般的情况下的证明, 作为我们给出证明的后续部分, 见理论题 2 和 3.) 我们需要以下引理.

引理 2.1 对于一个非负随机变量 Y, 有

$$E[Y] = \int_0^{\infty} P\{Y > y\}\mathrm{d}y$$

证明: 本证明中, 我们假定 Y 是一个连续型随机变量, 其密度函数为 f_Y. 此时有

$$\int_0^{\infty} P\{Y > y\}\,\mathrm{d}y = \int_0^{\infty}\int_y^{\infty} f_Y(x)\,\mathrm{d}x\,\mathrm{d}y$$

此处利用了事实 $P\{Y > y\} = \int_y^{\infty} f_Y(x)\,\mathrm{d}x$, 交换上式的积分次序, 可以得到

$$\int_0^{\infty} P\{Y > y\}\,\mathrm{d}y = \int_0^{\infty}\left(\int_0^x \mathrm{d}y\right) f_Y(x)\,\mathrm{d}x = \int_0^{\infty} x f_Y(x)\,\mathrm{d}x = E[Y] \qquad \square$$

命题 2.1 的证明: 对于任一函数 g, 其中 $g(x) \geqslant 0$, 根据引理 2.1, 有

$$E[g(X)] = \int_0^{\infty} P\{g(X) > y\}\,\mathrm{d}y = \int_0^{\infty}\int_{x:g(x)>y} f(x)\,\mathrm{d}x\,\mathrm{d}y$$
$$= \int_{x:g(x)>0}\int_0^{g(x)} \mathrm{d}y f(x)\,\mathrm{d}x = \int_{x:g(x)>0} g(x)f(x)\,\mathrm{d}x$$

这样, 命题就得到了证明. $\square$

例 2c 一根长度为 1 的棍子在点 U 处断开, 其中 U 服从 $(0,1)$ 上的均匀分布, 求包含点 p 的那一截的长度的期望值 $(0 \leqslant p \leqslant 1)$.

解: 令 $L_p(U)$ 表示包含 p 的那一截的长度, $L_p(U)$ 具有下列表达式 (见图 5.2

的图解)

$$L_p(U) = \begin{cases} 1-U & U<p \\ U & U>p \end{cases}$$

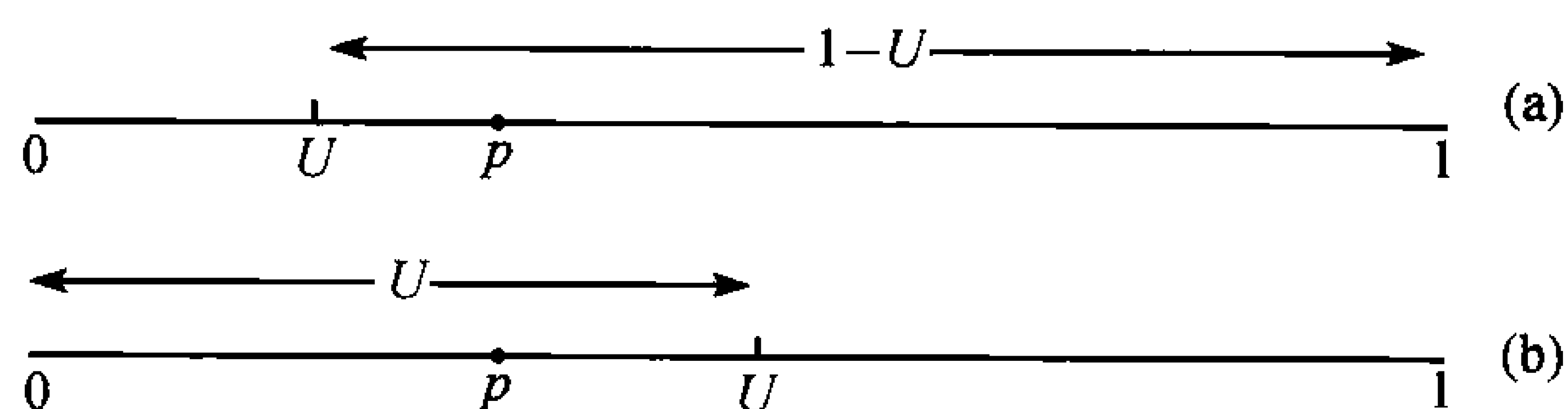


图 5.2 包含 p 点的部分: (a) $U<p$, (b) $U>p$.

因此, 利用命题 2.1 有

$$\begin{aligned} E[L_p(U)] &= \int_0^1 L_p(u)\mathrm{d}u = \int_0^p (1-u)\mathrm{d}u + \int_p^1 u\,\mathrm{d}u \\ &= \frac{1}{2} - \frac{(1-p)^2}{2} + \frac{1}{2} - \frac{p^2}{2} = \frac{1}{2} + p(1-p) \end{aligned}$$

既然 $p(1-p)$ 在 $p=1/2$ 时取最大值, 这样, 当 p 是棍子的中点时, 包含 p 点的那一截的长度的期望取得最大值, 这一点很有意思. ■

例 2d 假设你去赴约, 如果早到 s 分钟, 那么将要花费 cs 元, 如果晚到 s 分钟, 那么将要花费 ks 元. 又假设从你所在地点到约会地点路途所要花的时间为一个随机变量, 其概率密度函数为 f, 如果要使得花费的期望值最小, 你应该什么时候出发?

解: 令 X 表示路途所花时间, 如果你在约会前 t 分钟出发, 那么你的花费 $C_t(X)$ 的公式为

$$C_t(X) = \begin{cases} c(t-X) & \text{如果 } X \leqslant t \\ k(X-t) & \text{如果 } X \geqslant t \end{cases}$$

因此,

$$\begin{aligned} E[C_t(X)] &= \int_0^\infty C_t(x)f(x)\,\mathrm{d}x = \int_0^t c(t-x)f(x)\,\mathrm{d}x + \int_t^\infty k(x-t)f(x)\,\mathrm{d}x \\ &= ct\int_0^t f(x)\,\mathrm{d}x - c\int_0^t xf(x)\,\mathrm{d}x + k\int_t^\infty xf(x)\,\mathrm{d}x - kt\int_t^\infty f(x)\,\mathrm{d}x \end{aligned}$$

对上式求导可得

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}E[C_t(X)] &= ctf(t) + cF(t) - ctf(t) - ktf(t) + ktf(t) - k[1-F(t)] \\ &= (k+c)F(t) - k \end{aligned}$$

令上式等于 0, 便可得到, 在约会前 t^* 分钟出发可以使得花费的期望值最小, 其中

t^* 满足:

$$F(t^*) = \frac{k}{k+c}$$ ■

类似于第 4 章中的方法, 我们可以利用命题 2.1 证明以下推论.

推论 2.1 如果 a 和 b 都是常数, 那么 $E[aX+b] = aE[X]+b$

本推论的证明完全类似于离散型随机变量情形下的证明. 不同之处仅仅在于求和换成了积分, 分布列换成了密度函数.

连续型随机变量的方差的定义也同离散型情况是一样的, 也即, 如果 X 是一个连续随机变量, 期望值为 μ, 那么 X (任何类型的随机变量) 的方差定义如下:

$$\mathrm{Var}(X) = E[(X-\mu)^2]$$

另外一个公式就是,

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2$$

该公式的证明方法同离散型情形也是一样的.

例 2e 求例 2a 里的随机变量 X 的 $\mathrm{Var}(X)$.

解: 我们先来计算 $E[X^2]$,

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x = \int_0^1 2x^3\,\mathrm{d}x = \frac{1}{2}$$

由 $E[X] = 2/3$, 可以得到 $\mathrm{Var}(X) = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}$. ■

类似于离散型情形, 还可以证明: 对于常数 a 和 b, 有

$$\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X).$$

有几类比较重要的连续型随机变量, 它们在概率的应用中经常出现, 接下来的几节将要介绍它们.

5.3 均匀分布的随机变量

一个随机变量称为服从 $(0,1)$ 区间上的*均匀分布*(uniformly distirbuted), 如果它的密度函数为

$$f(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{其他} \end{cases} \tag{3.1}$$

显然, $f(x) \geqslant 0$, 且 $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = \int_0^1 \mathrm{d}x = 1$, 上式为一个概率密度函数. 因为仅当 $x \in (0,1)$ 时才有 $f(x) > 0$, 因此, X 必然取值在 $(0,1)$ 之间. 而且, 既然 $f(x)$ 对于任意 $x \in (0,1)$ 为常数, 则 X 在 $(0,1)$ 间任何值附近取值的概率都是相等的.

要证明这点, 注意到对于任意 $0 < a < b < 1$, 有

$$P\{a \leqslant X \leqslant b\} = \int_a^b f(x)\,\mathrm{d}x = b - a$$

也就是说, X 属于 $(0,1)$ 的任一子区间的概率等于该子区间的长度.

一般来说, 我们称 X 为区间 (α, β) 上服从均匀分布的随机变量, 如果它的密度函数如下

$$f(x) = \begin{cases} \dfrac{1}{\beta - \alpha} & \text{如果 } \alpha < x < \beta \\ 0 & \text{其他} \end{cases} \tag{3.2}$$

利用 $F(a) = \int_{-\infty}^{a} f(x)\,\mathrm{d}x$, 我们可由上式得到区间 (α, β) 上的均匀分布的随机变量的分布函数为

$$F(a) = \begin{cases} 0 & a \leqslant \alpha \\ \dfrac{a - \alpha}{\beta - \alpha} & \alpha < a < \beta \\ 1 & a \geqslant \beta \end{cases}$$

图 5.3 显示的就是 $f(a)$ 和 $F(a)$.

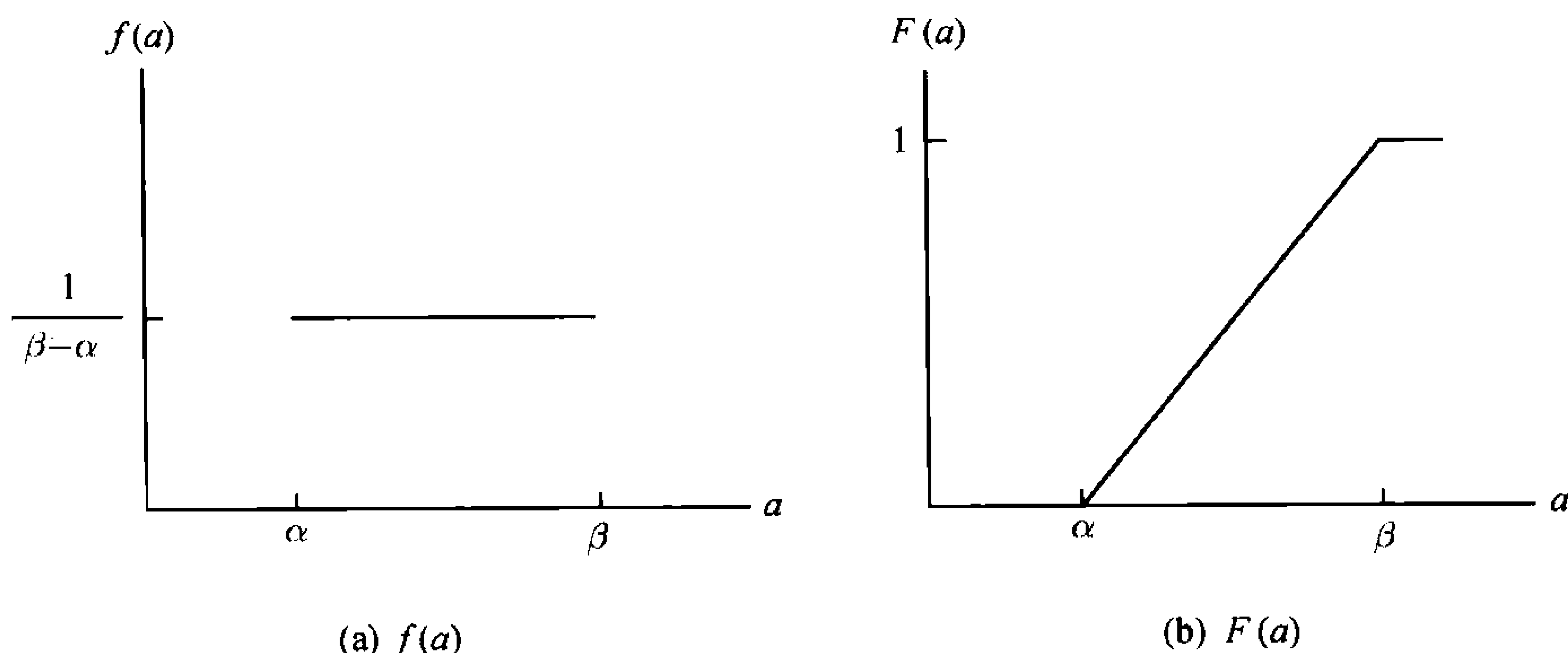


图 5.3　(α, β) 上均匀分布的密度函数 $f(a)$ 和分布函数 $F(a)$

例 3a　令 X 在 (α, β) 上服从均匀分布, 求 (a) $E[X]$; (b) $\mathrm{Var}(X)$.

解: (a)

$$E[X] = \int_{-\infty}^{\infty} x f(x)\mathrm{d}x = \int_{\alpha}^{\beta} \frac{x}{\beta - \alpha}\mathrm{d}x = \frac{\beta^2 - \alpha^2}{2(\beta - \alpha)} = \frac{\beta + \alpha}{2}$$

也就是说, 在某个区间上服从均匀分布的随机变量的期望等于该区间的中点的值.

(b) 为了计算 $\mathrm{Var}(X)$, 先计算 $E[X^2]$,

$$E[X^2] = \int_{\alpha}^{\beta} \frac{1}{\beta - \alpha} x^2 \mathrm{d}x = \frac{\beta^3 - \alpha^3}{3(\beta - \alpha)} = \frac{\beta^2 + \alpha\beta + \alpha^2}{3}$$

因此

$$\text{Var}(X) = \frac{\beta^2 + \alpha\beta + \alpha^2}{3} - \frac{(\alpha+\beta)^2}{4} = \frac{(\beta-\alpha)^2}{12}$$

这样, 在某个区间上服从均匀分布的随机变量的方差等于区间长度的平方除以 12. ■

例 3b 如果 X 服从 $(0, 10)$ 上的均匀分布, 计算如下概率:

(a) $X < 3$; (b) $X > 6$; (c) $3 < X < 8$.

解: (a) $P\{X < 3\} = \int_0^3 \frac{1}{10}\mathrm{d}x = \frac{3}{10}$

(b) $P\{X > 6\} = \int_6^{10} \frac{1}{10}\mathrm{d}x = \frac{4}{10}$ (c) $P\{3 < X < 8\} = \int_3^8 \frac{1}{10}\mathrm{d}x = \frac{1}{2}$ ■

例 3c 公共汽车从早上 7 点钟开始, 到达某一车站的时间间隔为 15 分钟. 也就是说, 汽车到达的时间为 7 点, 7 点 15, 7 点 30, 7 点 45, 等等. 如果某个乘客到达车站的时间服从 7 点到 7 点半之间的均匀分布, 求以下事件的概率:

(a) 他等公共汽车的时间不超过 5 分钟;

(b) 他等公共汽车的时间超过 10 分钟.

解: 令 X 表示从 7 点到该乘客到达车站的时间差 (分钟), 这样 X 就是一个区间在 (0, 30) 上服从均匀分布的随机变量. 乘客等待时间不超过 5 分钟, 当且仅当他到达时间为 7 点 10 分到 7 点 15 分之间, 或者 7 点 25 到 7 点 30 之间. 因此 (a) 所求概率为

$$P\{10 < X < 15\} + P\{25 < X < 30\} = \int_{10}^{15} \frac{1}{30}\mathrm{d}x + \int_{25}^{30} \frac{1}{30}\mathrm{d}x = \frac{1}{3}$$

类似地, 他等待时间超过 10 分钟, 当且仅当他到达时间为 7 点到 7 点 5 分之间, 或者 7 点 15 到 7 点 20 之间, 因此 (b) 所求概率为

$$P\{0 < X < 5\} + P\{15 < X < 20\} = \frac{1}{3}$$ ■

接下来的例子是法国数学家贝特朗于 1889 年首先考虑的问题, 通常称为贝特朗悖论. 它说明, 早期概率论中的概率这个概念来源于几何概率.

例 3d 考虑随机地从圆中取一根弦, 该弦的的长度大于该圆内接正三角形的边长的概率是多大?

解: 上述问题其实是无解的, 因为随机取弦的定义并不明确. 下面将用两种不同的方法来重新阐述这个问题.

第一种方法: 弦的位置可由它到圆心的距离确定, 此距离的变化范围为 0 到 r, 其中 r 为圆的半径. 这样, 当弦与圆心的距离小于 $r/2$ 时, 弦长将大于圆内接等边三角形的边长. 因此, 假设随机地取弦意味着弦到圆心的距离 D 服从 0 到 r 的均匀分布, 因此, 该弦的长度大于内接等边三角形的边长的概率为

$$P\left\{D < \frac{r}{2}\right\} = \frac{r/2}{r} = \frac{1}{2}$$

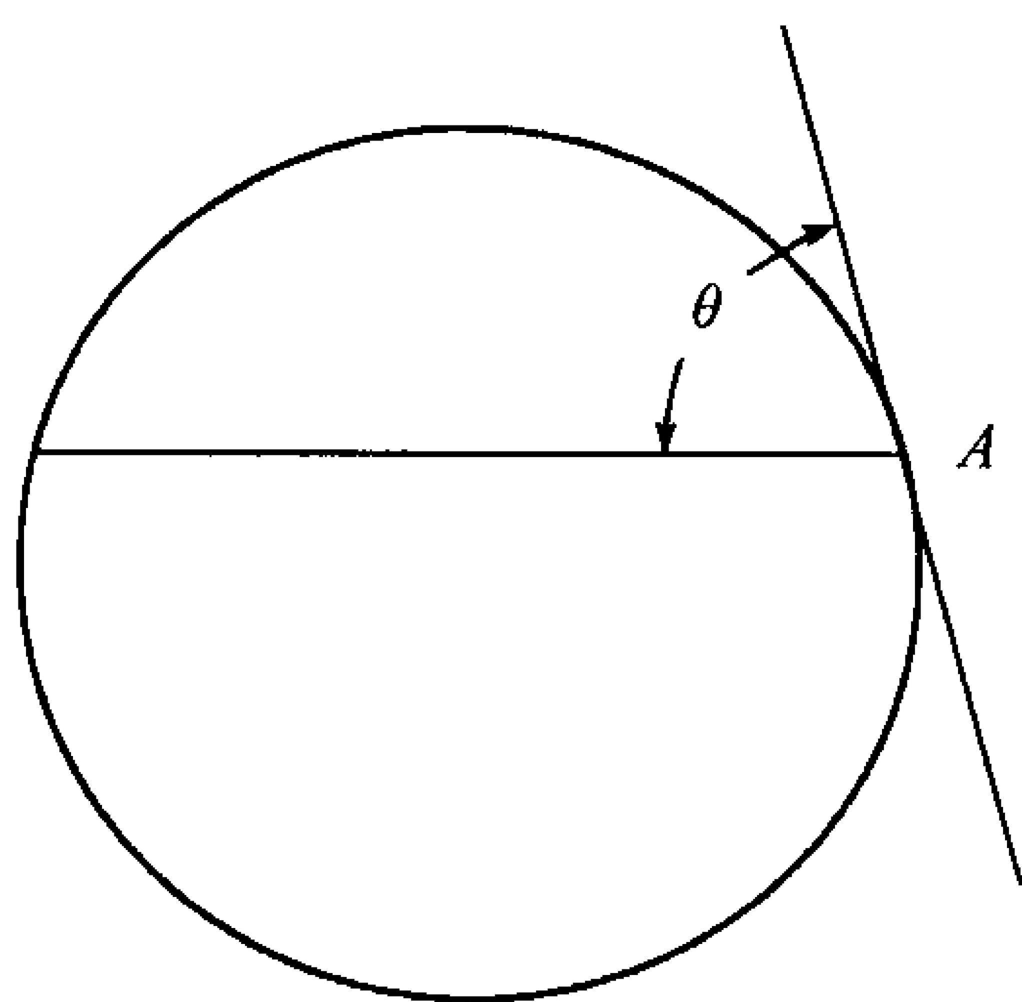


图 5.4 随机取弦

第二种方法是这样考虑随机取弦：通过弦的一端作切线，那么弦与切线之间的夹角 θ 的变化范围为 0° 到 180°, 它决定了弦的位置 (见图 5.4). 而且, 当 θ 在 60° 到 120° 之间时, 弦的长度大于圆内接等边三角形的边长. 因此, 假设随机的弦意味着 θ 在 0° 到 180° 之间均匀分布, 因此在这种假定下所求概率为

$$P\{60 < \theta < 120\} = \frac{120-60}{180} = \frac{1}{3}$$

注意到进行这样的随机试验时, 正确答案可能是 1/2, 也可能是 1/3. 假设桌上画了好多平行线, 平行线之间的距离为 $2r$, 现将一个直径为 $2r$ 的圆盘往桌上扔, 那么, 这个圆盘必定与某条平行线相交, 该平行线与圆盘相交形成一条弦. 这条弦的长度决定于圆盘的圆心在平面上的位置. 此时, 这条弦的长度分布与第一种情况相适应. 因此, 弦的长度大于内接正三角形边长的概率为 1/2. 如果在桌子上画一个半径为 r 的圆, 在圆周上取一点, 记为 A, 在 A 点上钉一根可以任意绕 A 点转动的针, 这个针与圆周总会相交而得到一条弦, 而这条弦的长度分布就与第二种情况相同, 其长度大于内接正三角形边长的概率为 1/3. ■

5.4 正态随机变量

我们称 X 为服从参数为 μ 和 σ^2 正态分布的随机变量, 或者就简单称为正态随机变量, 如果 X 的密度函数为

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} \qquad -\infty < x < \infty$$

该密度函数是一条关于 μ 对称的钟型曲线 (如图 5.5).

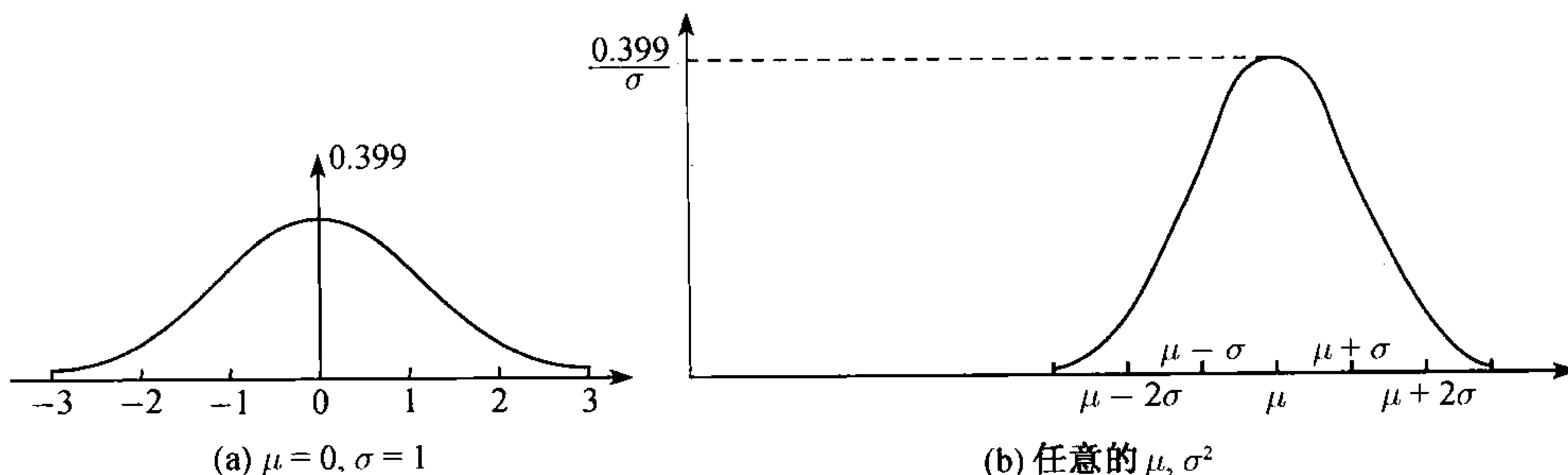


图 5.5 正态分布密度函数

1733 年, 法国数学家亚伯拉罕 · 棣莫弗引入了正态分布, 并将它用在近似计算二项随机变量当 n 很大时的取值概率. 这个结果后来被拉普拉斯和其他一些数学家做了推广, 现在被概括为一条概率论定理 ——中心极限定理. 这点将要在第 8 章介绍. 中心极限定理是概率论中两个重要的结果之一[1], 它给出了实际中出现的许多随机变量服从正态分布的理论根据. 下列都是一些服从正态分布的例子: 一个成人的身高, 一个气体分子在任意给定方向上的运动速度, 测量物体质量时的误差等.

为了证明 $f(x)$ 的确是一个密度函数, 我们要证明

$$\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}\mathrm{e}^{-(x-\mu)^2/2\sigma^2}\mathrm{d}x=1$$

作替换 $y=(x-\mu)/\sigma$ 可得

$$\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}\mathrm{e}^{-(x-\mu)^2/2\sigma^2}\mathrm{d}x=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\mathrm{e}^{-y^2/2}\mathrm{d}y$$

因此, 我们必需证明

$$\int_{-\infty}^{\infty}\mathrm{e}^{-y^2/2}\,\mathrm{d}y=\sqrt{2\pi}$$

令 $I=\int_{-\infty}^{\infty}\mathrm{e}^{-y^2/2}\mathrm{d}y$, 那么

$$I^2=\int_{-\infty}^{\infty}\mathrm{e}^{-y^2/2}\mathrm{d}y\int_{-\infty}^{\infty}\mathrm{e}^{-x^2/2}\mathrm{d}x=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathrm{e}^{-(y^2+x^2)/2}\,\mathrm{d}y\,\mathrm{d}x$$

我们通过将变量转换为极坐标形式来求解上面的二重积分 (也即令 $x=r\cos\theta, y=r\sin\theta$, 且 $\mathrm{d}y\,\mathrm{d}x=r\,\mathrm{d}\theta\,\mathrm{d}r$),

$$I^2=\int_0^{\infty}\int_0^{2\pi}\mathrm{e}^{-r^2/2}r\,\mathrm{d}\theta\mathrm{d}r=2\pi\int_0^{\infty}r\mathrm{e}^{-r^2/2}\,\mathrm{d}r=-2\pi\mathrm{e}^{-r^2/2}\Big|_0^{\infty}=2\pi$$

因此, $I=\sqrt{2\pi}$, 这样结论就得到了证明.

另外一个关于正态分布的重要结论是: 如果 X 是一个服从参数为 μ 和 σ 的正态分布的随机变量, 那么 $Y=aX+b$ 也服从正态分布, 其参数为 $a\mu+b$ 和 $a^2\sigma^2$. 为了证明这点, 假设 $a>0$($a<0$ 时证明类似). 令 F_Y 表示 Y 的分布函数, 那么

$$F_Y(x)=P\{Y\leqslant x\}=P\{aX+b\leqslant x\}=P\Big\{X\leqslant\frac{x-b}{a}\Big\}=F_X\Big(\frac{x-b}{a}\Big)$$

其中 F_X 为 X 的分布函数. 求导可得 Y 的密度函数为

$$\begin{aligned}f_Y(x)&=\frac{1}{a}f_X\Big(\frac{x-b}{a}\Big)=\frac{1}{\sqrt{2\pi}a\sigma}\exp\Big\{-\Big(\frac{x-b}{a}-\mu\Big)^2/2\sigma^2\Big\}\\&=\frac{1}{\sqrt{2\pi}a\sigma}\exp\{-(x-b-a\mu)^2/2(a\sigma)^2\}\end{aligned}$$

---

① 另一个是强大数定律.

这样就证明了, Y 确实服从参数为 $a\mu+b$ 和 $a^2\sigma^2$ 的正态分布.

上述结论的一个重要应用就是, 如果 X 是一个参数为 (μ, σ^2) 的正态分布随机变量, 那么 $Z=(X-\mu)/\sigma$ 就是一个服从参数为 $(0, 1)$ 的正态分布. 这样的随机变量称为标准正态随机变量.

接下来证明, 正态分布的参数 μ 和 σ^2 分别代表了它的期望和方差.

例 4a　X 是一服从参数 μ 和 σ^2 的正态分布的随机变量, 求 $E[X]$ 和 $\mathrm{Var}(X)$.

解: 先从计算标准正态分布随机变量 $Z=(X-\mu)/\sigma$ 的期望和方差开始, 有

$$E[Z]=\int_{-\infty}^{\infty} x f_Z(x)\mathrm{d}x=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x\mathrm{e}^{-x^2/2}\,\mathrm{d}x=-\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2}\Big|_{-\infty}^{\infty}=0$$

因此,

$$\mathrm{Var}(Z)=E[Z^2]=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x^2\mathrm{e}^{-x^2/2}\,\mathrm{d}x$$

通过分部积分 $(u=x, \mathrm{d}v=x\mathrm{e}^{-x^2/2})$ 得到

$$\mathrm{Var}(Z)=\frac{1}{\sqrt{2\pi}}\Big(-x\mathrm{e}^{-x^2/2}\Big|_{-\infty}^{\infty}+\int_{-\infty}^{\infty}\mathrm{e}^{-x^2/2}\,\mathrm{d}x\Big)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\mathrm{e}^{-x^2/2}\,\mathrm{d}x=1$$

由 $X=\mu+\sigma Z$ 得到

$$E[X]=\mu+\sigma E[Z]=\mu \quad 及 \quad \mathrm{Var}(X)=\sigma^2\mathrm{Var}(Z)=\sigma^2$$ ■

按照传统的记法, 一般将标准正态分布的分布函数记为 $\Phi(x)$, 也即

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-y^2/2}\,\mathrm{d}y$$

对于一个非负数 x, $\Phi(x)$ 的值在表 5.1 已经给出. 对于一个负数 x, $\Phi(x)$ 的值可以通过 (4.1) 式计算得到

$$\Phi(-x)=1-\Phi(x)\qquad -\infty<x<\infty \tag{4.1}$$

公式 (4.1), 可以利用标准正态密度函数的对称性得到, 证明留作习题. 该公式表明, 如果 Z 是一个标准正态分布的随机变量, 那么

$$P\{Z\leqslant -x\}=P\{Z>x\}\qquad -\infty<x<\infty$$

因为当 X 服从参数为 μ 和 σ^2 的正态分布时, $Z=(X-\mu)/\sigma$ 服从标准正态分布, 因此, X 的分布函数可以写成:

$$F_X(a)=P\{X\leqslant a\}=P\Big(\frac{X-\mu}{\sigma}\leqslant\frac{a-\mu}{\sigma}\Big)=\Phi\Big(\frac{a-\mu}{\sigma}\Big)$$

例 4b　如果 X 服从正态分布, 参数为 $\mu=3$ 和 $\sigma^2=9$, 求

(a) $P\{2<X<5\}$;　(b) $P\{X>0\}$;　(c) $P\{|X-3|>6\}$.

表 5.1 $\Phi(x)$: 标准正态分布密度曲线下 x 左侧的面积

x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

解: (a) $P\{2 < X < 5\} = P\left\{\frac{2-3}{3} < \frac{X-3}{3} < \frac{5-3}{3}\right\} = P\left\{-\frac{1}{3} < Z < \frac{2}{3}\right\}$

$$= \Phi\left(\frac{2}{3}\right) - \Phi\left(-\frac{1}{3}\right) = \Phi\left(\frac{2}{3}\right) - \left[1 - \Phi\left(\frac{1}{3}\right)\right] \approx 0.3779$$

(b) $P\{X > 0\} = P\left\{\frac{X-3}{3} > \frac{0-3}{3}\right\} = P\{Z > -1\} = 1 - \Phi(-1) = \Phi(1) \approx 0.8413$

(c) $P\{|X-3| > 6\} = P\{X > 9\} + P\{X < -3\}$

$$= P\left\{\frac{X-3}{3} > \frac{9-3}{3}\right\} + P\left\{\frac{X-3}{3} < \frac{-3-3}{3}\right\}$$

$$= P\{Z > 2\} + P\{Z < -2\}$$

$$= 1 - \Phi(2) + \Phi(-2) = 2[1 - \Phi(2)] \approx 0.0456$$ ■

例 4c 进行一次考试, 如果所有考生所得的分数可近似地表示为正态密度函数 (换句话说, 各级考分的频率图近似地呈现正态密度的钟形曲线.), 则通常认为这次考试 (就合理地划分考生成绩的等级而言) 是可取的. 教师经常用考试的分数去估计正态参数 μ 与 σ^2, 然后把分数超过 $\mu+\sigma$ 的评为 A 等, 分数在 μ 到 $\mu+\sigma$ 之间的评为 B 等, 分数在 $\mu-\sigma$ 到 μ 之间的评为 C 等, 分数在 $\mu-2\sigma$ 到 $\mu-\sigma$ 之间评为 D 等, 分数在 $\mu-2\sigma$ 以下者评为 F 等. (有时称这种方法为"曲线上"划分等级法) 由于

$$P\{X > \mu+\sigma\} = P\left\{\frac{X-\mu}{\sigma} > 1\right\} = 1 - \Phi(1) \approx 0.1587$$

$$P\{\mu < X < \mu+\sigma\} = P\left\{0 < \frac{X-\mu}{\sigma} < 1\right\} = \Phi(1) - \Phi(0) \approx 0.3413$$

$$P\{\mu-\sigma < X < \mu\} = P\left\{-1 < \frac{X-\mu}{\sigma} < 0\right\} = \Phi(0) - \Phi(-1) \approx 0.3413$$

$$P\{\mu-2\sigma < X < \mu-\sigma\} = P\left\{-2 < \frac{X-\mu}{\sigma} < -1\right\} = \Phi(2) - \Phi(1) \approx 0.1359$$

$$P\{X < \mu-2\sigma\} = P\left\{\frac{X-\mu}{\sigma} < -2\right\} = \Phi(-2) \approx 0.0228$$

所以, 近似地说, 这次考试中, 能获得 A 等的占 16%, B 等的占 34%, C 等的占 34%, D 等的占 14%, 成绩很差的占 2%. ■

例 4d 某人被指控为一个新生儿的父亲. 此案鉴定人作证时指出: 母亲的怀孕期 (即从受孕到婴儿出生的时间) 的天数近似地服从正态分布, 其参数为 $\mu = 270$, $\sigma^2 = 100$. 被告提供的证词表明, 他在孩子出生前 290 天出国, 而于出生前 240 天才回来. 如果被告事实上是这孩子的父亲, 试问那位母亲确有与证词相符的、过长或过短的怀孕期的概率是多少?

解: 设 X 表示怀孕期的天数, 并假定被告是这孩子的父亲, 那么孩子生于与证词相符的时间内的概率是

$$\begin{aligned} P\{X > 290 \text{或} X < 240\} &= P\{X > 290\} + P\{X < 240\} \\ &= P\left\{\frac{X-270}{10} > 2\right\} + P\left\{\frac{X-270}{10} < -3\right\} \\ &= 1 - \Phi(2) + 1 - \Phi(3) \approx 0.0241 \end{aligned}$$ ■

例 4e 考虑从 A 地到 B 地通过电信传送一个二值信号, 0 或 1. 然而, 数据通过电信传送过程中会遇到噪声干扰. 为了减少传送出错的概率, 当传送的信息为 1 时, 将传送值 2, 传送信息是 0 时, 就传送值 -2. 如果 $x, x = \pm 2$ 为在 A 地传送的数值, R 为在 B 地接收到的数值, ($R = x + N$, N 为噪声干扰), 当信号在 B 处接收后, 按如下解码规则.

如果 $R \geqslant 0.5$, 则认为是 1;　如果 $R < 0.5$, 则认为是 0.

如果噪声服从正态分布, 我们将要计算 N 为标准正态随机变量情形下的出错概率.

共有两类错误. 其一是信息 1 被错误地认为是 0; 另一类是信息 0 被错误地认为是 1. 第一类错误会在下列情形发生: 如果信息是 1, 且 $2+N < 0.5$, 而第二类错误会在下列情形发生: 信息是 0, 且 $-2+N \geqslant 0.5$. 因此,

$$P\{\text{错误}|\text{信息是1}\} = P\{N < -1.5\} = 1 - \Phi(1.5) \approx 0.0668$$

$$P\{\text{错误}|\text{信息是0}\} = P\{N \geqslant 2.5\} = 1 - \Phi(2.5) \approx 0.0062$$ ■

二项分布的正态近似

概率论里一个重要结论, 称为棣莫弗–拉普拉斯极限定理. 说明当 n 很大时, 参数为 (n,p) 的二项随机变量可用正态随机变量来近似. 正态分布的期望与方差与二项随机变量的期望和方差相同 $(np, np(1-p))$. 棣莫弗在 1733 年证明了 $p=1/2$ 的特殊情形. 尔后, 在 1812 年, 拉普拉斯对一般的 p 进行了证明. 更一般的叙述是: 我们可以如下将二项随机变量标准化, 先减去其均值 np, 然后再除以标准差 $\sqrt{np(1-p)}$, 那么经过标准化的随机变量 (均值为 0, 方差为 1) 的分布函数当 $n \to \infty$ 时收敛到标准正态分布函数.

> **棣莫弗 – 拉普拉斯极限定理**　在 n 次独立重复试验中, 设每次成功的概率为 p, 记成功的次数为 S_n, 则对任何 $a < b$ 有: 当 $n \to \infty$ 时
>
> $$P\left\{a \leqslant \frac{S_n - np}{\sqrt{np(1-p)}} \leqslant b\right\} \to \Phi(b) - \Phi(a)$$

由于上述定理仅是第 8 章研究的中心极限定理的一个特殊情形, 故这里不再证明.

大家看到, 对于二项分布, 我们已经有了两个可能的近似: 当 n 较大而 p 较小时, 泊松近似是一个很好的近似; 另外, 可以证明, 当 $np(1-p)$ 较大时, 正态近似相当好, 见图 5.6. [一般来说, 当 $np(1-p) \geqslant 10$ 时, 正态近似就相当好.]

例 4f　以 X 表示抛 40 次均匀硬币出现正面的次数. 试求 $X=20$ 的概率. 运用正态近似法, 再与精确解比较.

解: 为了利用正态近似, 注意到因为二项分布是离散整数值随机变量, 而正态分布为连续型随机变量, 因此最好在正态近似前将 $P\{X=i\}$ 写为 $P\{i-1/2 < X < i+1/2\}$[这也称为**连续性修正**(continuity correction)]. 这样

$$\begin{aligned} P\{X=20\} &= P\{19.5 \leqslant X < 20.5\} \\ &= P\left\{\frac{19.5-20}{\sqrt{10}} < \frac{X-20}{\sqrt{10}} < \frac{20.5-20}{\sqrt{10}}\right\} \end{aligned}$$

$$\approx P\left\{-0.16 < \frac{X-20}{\sqrt{10}} < 0.16\right\} \approx \Phi(0.16) - \Phi(-0.16) \approx 0.1272$$

而精确解为

$$P\{X=20\} = \binom{40}{20}\left(\frac{1}{2}\right)^{40} \approx 0.1254$$

■

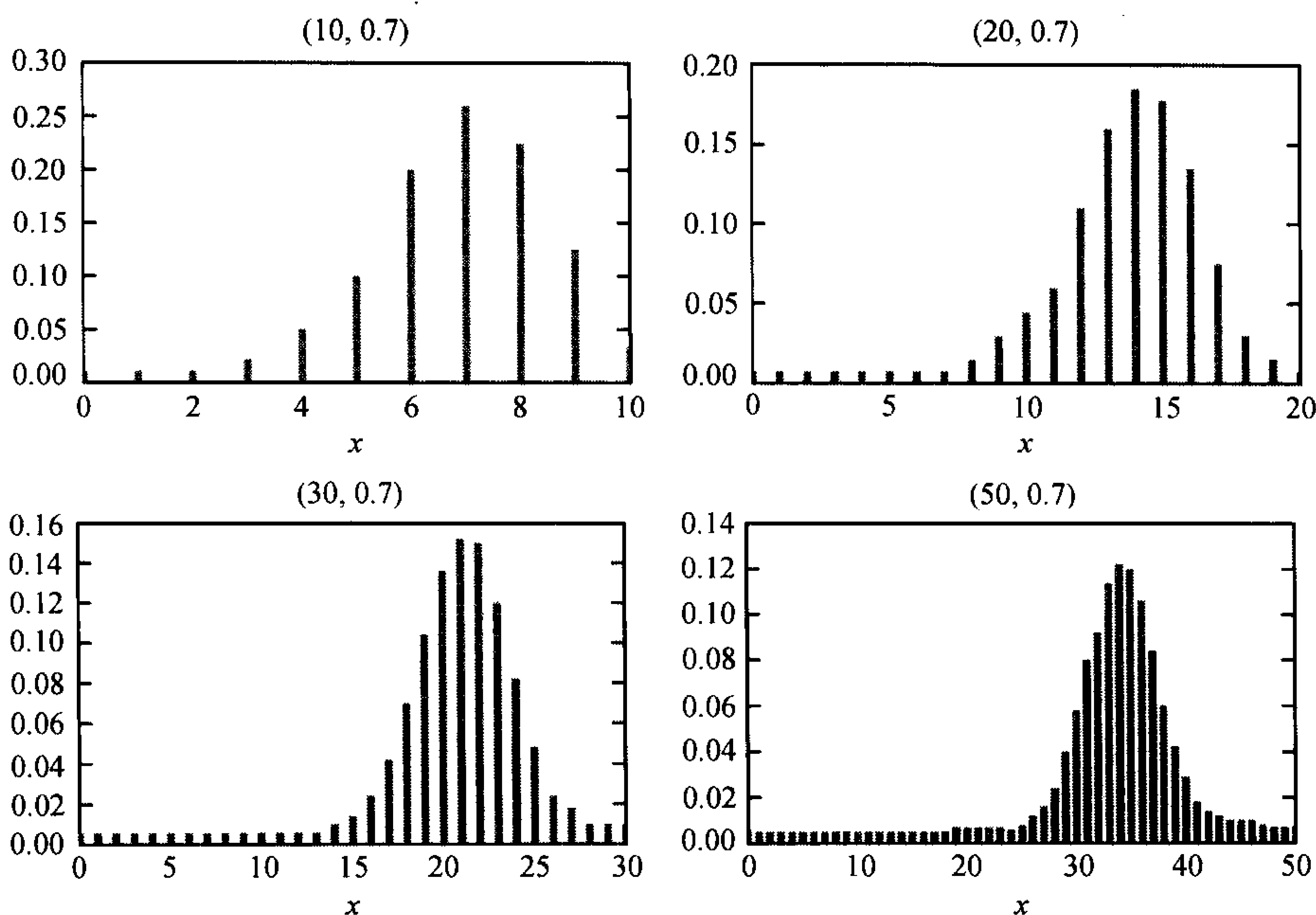


图 5.6　二项分布 (n, p) 的概率分布列随着 n 的增大越来越趋于正态分布

例 4g　某学院计划招收 150 名一年级新生. 根据以往的经验, 接到录取通知的人当中, 平均只有 30% 的人报到入学, 故学院给 450 名学生发录取通知书. 试求这所学院入学新生超过 150 名的概率.

解: 记 X 为入学新生人数, 那么 X 为以 $n=450, p=0.3$ 为参数的二项随机变量. 利用连续性修正及正态近似, 可得

$$P\{X \geqslant 150.5\} = P\left\{\frac{X-450\times 0.3}{\sqrt{450\times 0.3\times 0.7}} \geqslant \frac{150.5-450\times 0.3}{\sqrt{450\times 0.3\times 0.7}}\right\} \approx 1-\Phi(1.59) \approx 0.0559$$

这样, 在接到录取通知的人当中, 入学者超过 150 名的可能性不超过 6%. (这里我们做了怎样的独立性假设?)　■

例 4h　为测定能降低血液中胆固醇含量的某种食品的有效性, 营养学家让 100 个人吃这种食品. 经充分长的时间后, 化验他们的胆固醇含量. 如果至少有 65% 的人在吃了这种食物以后胆固醇含量降低, 则进行这项试验的营养学家就决定承认这种食品. 如果这种食品事实上对胆固醇含量不起作用, 试问这位营养学家承认它的

概率是多大?

解: 我们假定这种食品对降低胆固醇含量不起作用, 而一个人在吃了这种食品后碰巧胆固醇降低的概率为 1/2. 这样, 若以 X 表示胆固醇降低了的人数, 则当这种食品不影响胆固醇含量时, 营养学家承认它的概率是

$$\sum_{i=65}^{100}\binom{100}{i}\left(\frac{1}{2}\right)^{100} = P\{X \geqslant 64.5\} = P\left\{\frac{X - 100\times\frac{1}{2}}{\sqrt{100\times\frac{1}{2}\times\frac{1}{2}}} \geqslant 2.9\right\}$$
$$\approx 1 - \Phi(2.9) \approx 0.0019$$ ■

例 4i 纽约市有 52% 的居民支持禁止公共场合吸烟. 随机抽取 n 个纽约市民, 问支持这项禁令的人数超过 50% 的概率有多大, 如果

(a) $n = 11$; (b) $n = 101$; (c) $n = 1001$.

要想该概率值超过 0.95, n 需要多大?

解: 令 N 表示纽约市居民人数. 要回答上述问题, 我们必须首先理解样本大小为 n 的随机抽样就是从 N 个人当中按如下方式抽取 n 人: 使得 $\binom{N}{n}$ 种 n 个人的子集被抽到的可能性都是一样的. 这样, 记 S_n 为样本里支持禁令的人数, 它是一个超几何随机变量. 也即, S_n 的分布与从装有 N 个球的坛子里抽取 n 个球, 其中取出的白球数的分布是一样的 (其中 $0.52N$ 的球为白球). 但因为 N 和 $0.52N$ 对于 n 来说都是很大的数, 根据二项分布对超几何分布的近似 (见 4.8.3 节) 可知, 也即 S_n 的分布与参数为 n 和 $p = 0.52$ 的二项分布是很接近的. 再利用正态分布对二项分布的近似可得

$$P\{S_n > 0.5n\} = P\left\{\frac{S_n - 0.52n}{\sqrt{n\times 0.52\times 0.48}} > \frac{0.5n - 0.52n}{\sqrt{n\times 0.52\times 0.48}}\right\}$$
$$= P\left\{\frac{S_n - 0.52n}{\sqrt{n\times 0.52\times 0.48}} > -0.04\sqrt{n}\right\} \approx \Phi(0.04\sqrt{n})$$

因此,

$$P\{S_n > 0.5n\} \approx \begin{cases} \Phi(0.1328) = 0.5528 & \text{如果 } n = 11 \\ \Phi(0.4020) = 0.6562 & \text{如果 } n = 101 \\ \Phi(1.2665) = 0.8973 & \text{如果 } n = 1001 \end{cases}$$

为了使得该概率大于 0.95, 我们需要 $\Phi(0.04\sqrt{n}) > 0.95$. 因为 $\Phi(x)$ 为一单调递增函数, 且 $\Phi(1.645) = 0.95$, 因此 $0.04\sqrt{n} > 1.645$, 也即 $n \geqslant 1691.266$. 因此, 样本大小至少为 1692. ■

关于正态分布的历史注记

正态分布是法国数学家亚伯拉罕 · 棣莫弗在 1733 年引入的. 他利用正态分布

求出了有关抛掷硬币试验中随机事件的概率的近似值. 当时称正态分布为指数钟形曲线. 1809 年, 德国著名数学家高斯以正态分布作为主要工具预测天文学中星体的位置, 这时才展现了正态分布的应用价值. 此后, 正态分布就称为高斯分布.

在十九世纪后半叶, 大部分统计学家认为大部分数据的直方图都具有高斯钟形曲线的形状. 事实上, 大家认为正常的数据集合应该具有这种形状. 由英国统计学家卡尔 · 皮尔森开始, 将高斯曲线简称为正态曲线. (中心极限定理为许多数据具有正态分布的事实提供了合理的解释, 我们将在第 8 章讲述这一定理.)

亚伯拉罕 · 棣莫弗 (1667—1754)

现在统计学已经普及, 统计学家具有很好的工作环境. 然而, 统计学的诞生地却是在 18 世纪初的伦敦, 一所黑暗的、肮脏的赌窟, 称为屠夫咖啡屋的地方. 亚伯拉罕 · 棣莫弗是一个来自天主教法国的耶稣教难民, 为生计, 他要为各种赌博计算赔钱的概率.

虽然亚伯拉罕 · 棣莫弗在咖啡屋内谋生存, 但他是一位著名的数学家, 他发现了正态曲线. 他还是皇家学会的会员, 并且是著名科学家牛顿的朋友.

统计学家卡尔 · 皮尔森想象棣莫弗在屠夫咖啡屋内工作的情景:"我想象棣莫弗坐在咖啡屋内肮脏的小桌边, 旁边坐着一位破产的赌徒. 而牛顿从嘈杂的人群走向棣莫弗的小桌边, 拉出他的朋友. 在艺术家的想象中, 这是一幅多么伟大的艺术杰作啊。"

卡尔 · 弗里德里克 · 高斯 (1777—1855)

高斯, 正态曲线的最早应用者之一, 是一位伟大的数学家. 著名的数学史学家 E. T. Bell 在 1954 年的著作《数学人物》(*Men of Mathematics*) 中, 有一章名为"数学王子"中提到了三位数学家, 阿基米德、牛顿和高斯. "他们三位是在最伟大数学家之列, 我们不可以以通常眼光来评价他们的贡献的大小. 他们在纯数学和应用数学领域内推波助澜. 阿基米德特别推崇纯数学, 牛顿恰恰是把他的数学发现应用于科学研究, 而高斯宣称无论是纯数学还是应用数学对他而言都是一样的. "

5.5 指数随机变量

如果随机变量的密度函数如下: 对于 $\lambda > 0$, 有

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{当} x \geqslant 0 \\ 0 & \text{当} x < 0 \end{cases}$$

那么称该随机变量为参数为 λ 的指数分布的随机变量 (或简称为指数随机变量). 指数随机变量的分布函数如下:

$$F(a) = P\{X \leqslant a\} = \int_0^a \lambda e^{-\lambda x} dx = -e^{-\lambda x}\Big|_0^a = 1 - e^{-\lambda a} \qquad a \geqslant 0$$

下面将要说明参数 λ 就等于期望值的倒数.

例 5a 令 X 为一参数为 λ 的指数随机变量, 计算 (a) $E[X]$, (b) $\mathrm{Var}(X)$.

解: (a) 因为密度函数为

$$f(x)=\begin{cases}\lambda\mathrm{e}^{-\lambda x} & x\geqslant 0\\ 0 & x<0\end{cases}$$

因此, 对于 $n>0$, 有

$$E[X^n]=\int_0^\infty x^n\lambda\mathrm{e}^{-\lambda x}\mathrm{d}x$$

分部积分 ($\lambda\mathrm{e}^{-\lambda x}\mathrm{d}x=dv, u=x^n$), 可以得到

$$\begin{aligned}E[X^n]&=-x^n\mathrm{e}^{-\lambda x}\big|_0^\infty+\int_0^\infty \mathrm{e}^{-\lambda x}nx^{n-1}\mathrm{d}x\\&=0+\frac{n}{\lambda}\int_0^\infty \lambda\mathrm{e}^{-\lambda x}x^{n-1}\mathrm{d}x=\frac{n}{\lambda}E[X^{n-1}]\end{aligned}$$

令 $n=1$, 及 $n=2$, 可以得到

$$E[X]=\frac{1}{\lambda}\qquad E[X^2]=\frac{2}{\lambda}E[X]=\frac{2}{\lambda^2}$$

(b) 因此

$$\mathrm{Var}(X)=\frac{2}{\lambda^2}-\left(\frac{1}{\lambda}\right)^2=\frac{1}{\lambda^2}$$

即指数分布的期望值等于参数 λ 的倒数, 而方差等于期望的平方. ■

在实践中, 指数分布经常作为某个事件发生的等待时间的分布而出现. 比如, 地震发生的时间间隔 (从现在开始计算), 一场新的战争爆发时间间隔, 从现在开始到你接到一个误拨的电话的时间间隔, 等等, 这些都是实践中的指数随机变量. (关于这种现象的理论解释可参考 4.7 节.)

例 5b 假设某个电话的通话时长 (单位: 分钟) 为参数为 $\lambda=1/10$ 的指数随机变量. 假设某人正好在你之前到达电话亭, 求以下事件概率:

(a) 你的等待时间超过 10 分钟; (b) 你的等待时间在 10 到 20 分钟之间.

解: 令 X 表示该人通话时长, 那么所求概率为:

(a) $P\{X>10\}=1-F(10)=\mathrm{e}^{-1}\approx 0.368$

(b) $P\{10<X<20\}=F(20)-F(10)=\mathrm{e}^{-1}-\mathrm{e}^{-2}\approx 0.233$ ■

我们称一个非负随机变量 X 是**无记忆的**(memoryless), 如果

$$P\{X>s+t|X>t\}=P\{X>s\}\qquad \text{对所有的 } s,t\geqslant 0 \text{ 成立} \tag{5.1}$$

如果我们令 X 为某个设备的寿命, 上式说明了在已知该设备已经使用 t 小时的条件下寿命至少为 $s+t$ 的概率, 与开始时寿命至少为 s 小时的概率是一样的. 换句

话说, 如果该设备在使用 t 小时后还能使用, 那么"剩余的"寿命同一开始时的寿命的分布是一样的 (即就好像该设备对已经使用了 t 小时没有记忆似的).

条件 (5.1) 又等价于

$$\frac{P\{X > s+t, X > t\}}{P\{X > t\}} = P\{X > s\}$$

或者

$$P\{X > s+t\} = P\{X > s\}P\{X > t\} \tag{5.2}$$

当 X 服从指数分布时, 公式 (5.2) 是成立的 ($\mathrm{e}^{-\lambda(s+t)} = \mathrm{e}^{-\lambda s}\mathrm{e}^{-\lambda t}$). 因此, 指数随机变量是无记忆的.

例 5c 某个邮局有两个职员, 假设当史密斯先生走进邮局时, 他发现两个职员正在分别接待琼斯女士和布朗先生. 史密斯先生被告知, 一旦处理完琼斯或者布朗的事情, 就立即接待他. 如果职员给每位顾客的服务时长都服从参数为 λ 的指数分布, 那么三人中, 史密斯先生是最后一个办完事情的概率是多大?

解: 从史密斯先生开始接受服务时考虑, 此时, 琼斯女士和布朗先生中有一个已经离开, 另一个仍在继续接受服务. 然而, 因为指数分布是没有记忆的, 因此, 仍在接受服务的顾客 (琼斯女士或布朗先生) 的办事时长服从参数为 λ 的指数分布, 这与此时刚开始服务是一样的. 因此, 由对称性可得, 剩下的一个人在史密斯先生之前完成服务的概率为 1/2. ■

可以证明, 不仅是指数分布具有无记忆性, 而且指数分布是唯一具有无记忆性的分布. 为了证明这点, 假设 X 是无记忆性的, 且令 $\bar{F}(x) = P\{X > x\}$. 那么, 利用公式 (5.2), 可以得到 $\bar{F}(s+t) = \bar{F}(s)\bar{F}(t)$, 即 $\bar{F}(\cdot)$ 满足函数方程 $g(s+t) = g(s)g(t)$. 然而, 已经证明该函数方程的唯一的非平凡右连续解就是[①]

$$g(x) = \mathrm{e}^{-\lambda x} \tag{5.3}$$

又因为分布函数总是右连续的, 因此我们有

$$\bar{F}(x) = \mathrm{e}^{-\lambda x} \quad \text{或者} \quad F(x) = P\{X \leqslant x\} = 1 - \mathrm{e}^{-\lambda x}$$

---

① 可以如下证明 (5.3) 式: 如果 $g(s+t) = g(s)g(t)$, 那么

$$g\left(\frac{2}{n}\right) = g\left(\frac{1}{n} + \frac{1}{n}\right) = g^2\left(\frac{1}{n}\right)$$

重复以上计算可以得到 $g(m/n) = g^m(1/n)$. 而且,

$$g(1) = g\left(\frac{1}{n} + \frac{1}{n} + \cdots + \frac{1}{n}\right) = g^n\left(\frac{1}{n}\right) \quad \text{或} \quad g\left(\frac{1}{n}\right) = (g(1))^{1/n}$$

因此, $g(m/n) = (g(1))^{m/n}$. 因为 g 是右连续的, 可以得出 $g(x) = (g(1))^x$. 又因 $g(1) = \left(g(\frac{1}{2})\right)^2 \geqslant 0$, 我们可以得到 $g(x) = \mathrm{e}^{-\lambda x}$, 其中 $\lambda = -\ln(g(1))$.

这样, 就证明了 X 服从指数分布.

例 5d 假设汽车在电池用完之前跑的英里数服从均值为 10 000 英里的指数分布, 如果某人要计划开始一个 5000 英里的旅行, 那么他不用更换电池就能跑完全程的概率是多大? 如果不服从指数分布呢?

解: 由指数分布的无记忆性可以得到, 电池剩下的寿命 (以 1000 英里为单位) 服从参数为 $\lambda = 1/10$ 的指数分布. 因此, 所求概率为

$$P\{\text{剩余寿命} > 5\} = 1 - F(5) = \mathrm{e}^{-5\lambda} = \mathrm{e}^{-1/2} \approx 0.604$$

然而, 如果剩余寿命 F 的分布不是指数分布, 那么对应概率为

$$P\{\text{寿命} > t+5|\ \text{寿命} > t\} = \frac{1-F(t+5)}{1-F(t)}$$

其中 t 就是旅行前电池已经使用的寿命. 因此, 如果不是指数分布, 那么在计算所求概率之前还需要了解其他信息 (也即 t). ■

指数分布的一个变形是一种取值或正或负, 但是绝对值服从参数为 $\lambda(\lambda \geqslant 0)$ 的指数分布. 这样的随机变量也称为拉普拉斯随机变量①, 它的密度函数为

$$f(x) = \frac{1}{2}\lambda\mathrm{e}^{-\lambda|x|} \qquad -\infty < x < \infty$$

其分布函数如下:

$$F(x) = \begin{cases} \dfrac{1}{2}\displaystyle\int_{-\infty}^{x} \lambda\mathrm{e}^{\lambda x}\mathrm{d}x \\ \dfrac{1}{2}\displaystyle\int_{-\infty}^{0} \lambda\mathrm{e}^{\lambda x}\mathrm{d}x + \dfrac{1}{2}\displaystyle\int_{0}^{x} \lambda\mathrm{e}^{-\lambda x}\mathrm{d}x \end{cases} = \begin{cases} \dfrac{1}{2}\mathrm{e}^{\lambda x} & x < 0 \\ 1 - \dfrac{1}{2}\mathrm{e}^{-\lambda x} & x > 0 \end{cases}$$

例 5e 我们来重新考虑例 4e, 从 A 地传送一个二值信息到 B 地. 当信息为 1 时, 传送 2, 当信息为 0 时传送 -2. 然而, 通信噪声 N 不再是标准正态随机变量, 而是参数为 $\lambda = 1$ 的拉普拉斯随机变量. 假设如果在 B 地收到 R, 信息如下解码:

如果 $R \geqslant 0.5$, 那么认为是 1;

如果 $R < 0.5$, 那么认为是 0.

这种情形下, 如果噪声为参数 $\lambda = 1$ 的拉普拉斯随机变量, 那么两类错误的概率如下

$$P\{\text{错误}|\text{信息是}1\} = P\{N < -1.5\} = \frac{1}{2}\mathrm{e}^{-1.5} \approx 0.1116$$

$$P\{\text{错误}|\text{信息是}0\} = P\{N \geqslant 2.5\} = \frac{1}{2}\mathrm{e}^{-2.5} \approx 0.041$$

将此结论与例 4e 对比, 可以发现, 噪声为参数为 $\lambda = 1$ 的拉普拉斯随机变量时的错误概率要大于为正态随机变量时的概率. ■

① 有时也称为双指数型随机变量.

危险率函数

考虑一个正连续型随机变量 X, 我们将它解释为某个零件的寿命, 具有分布函数 F 以及分布密度 f. **危险率**(Hazard rate, 有时也称为失效率, failure rate)$\lambda(t)$ 定义如下:

$$\lambda(t) = \frac{f(t)}{\bar{F}(t)}, \qquad \text{其中 } \bar{F} = 1 - F$$

为了解释 $\lambda(t)$, 考虑该零件已经使用了 t 小时, 我们来求它不能继续使用 $\mathrm{d}t$ 小时的概率, 也即求 $P\{X \in (t, t+\mathrm{d}t)|X > t\}$. 利用条件概率公式,

$$P\{X \in (t, t+\mathrm{d}t)|X > t\} = \frac{P\{X \in (t, t+\mathrm{d}t), X > t\}}{P\{X > t\}} = \frac{P\{X \in (t, t+\mathrm{d}t)\}}{P\{X > t\}} \approx \frac{f(t)}{\bar{F}(t)}\mathrm{d}t$$

因此, $\lambda(t)$ 表示了"年龄"为 t 的零件不能继续使用的条件概率强度.

现在假设寿命服从指数分布, 那么, 利用它的无记忆性, 可以得到对于一个"年龄"为 t 的零件, 它剩下的生存时间同一个新零件是一样的. 因此, $\lambda(t)$ 必然是一个常数. 这点可以检验如下

$$\lambda(t) = \frac{f(t)}{\bar{F}(t)} = \frac{\lambda \mathrm{e}^{-\lambda t}}{\mathrm{e}^{-\lambda t}} = \lambda$$

因此, 指数分布的危险率函数是一个常数. 由危险率函数的定义可知, 由分布可以计算它的危险率函数. 反过来, 由危险率函数也可以唯一地确定它的分布函数. 为证明这一点, 将 $\lambda(t)$ 的公式

$$\lambda(t) = \frac{\frac{\mathrm{d}}{\mathrm{d}t}F(t)}{1 - F(t)}$$

两边积分可得

$$\ln(1 - F(t)) = -\int_0^t \lambda(t)\mathrm{d}t + k$$

或

$$1 - F(t) = \mathrm{e}^k \exp\left\{-\int_0^t \lambda(t)\mathrm{d}t\right\}$$

令 $t = 0$ 可得 $k = 0$, 因此

$$F(t) = 1 - \exp\left\{-\int_0^t \lambda(t)\mathrm{d}t\right\} \tag{5.4}$$

因此, 一个正连续型随机变量的分布函数可由其危险率函数确定. 比如说, 如果随机变量具有线性危险率函数, 即如果

$$\lambda(t) = a + bt$$

那么其分布函数为

$$F(t) = 1 - \mathrm{e}^{-at - bt^2/2}$$

求导可得

$$f(t) = (a + bt)\mathrm{e}^{-(at + bt^2/2)} \qquad t \geqslant 0$$

当 $a = 0$ 时, 上述即为熟知的瑞利 (Rayleigh) 分布的密度函数.

例 5f 人们经常听到, 各个年龄段吸烟者的死亡率是非吸烟者死亡率的 2 倍, 这是什么意思? 是不是说对于同年龄的非吸烟者和吸烟者来说, 前者活到一个给定时间的概率是后者的两倍?

解: 令 $\lambda_s(t)$ 表示年龄为 t 的吸烟者的危险率, 而 $\lambda_n(t)$ 表示年龄为 t 的非吸烟者的危险率, 上述的含义等价于 $\lambda_s(t)=2\lambda_n(t)$.
一个年龄为 A 的非吸烟者能活到年龄 B 的概率 $(A<B)$ 为

$$
\begin{aligned}
&P\{\text{年龄为 } A \text{ 的非吸烟者能活到年龄} B\}\\
&=P\{\text{非吸烟者的寿命} > B | \text{非吸烟者的寿命} > A\}=\frac{1-F_{\text{non}}(B)}{1-F_{\text{non}}(A)}\\
&=\frac{\exp\left\{-\int_0^B \lambda_n(t)\mathrm{d}t\right\}}{\exp\left\{-\int_0^A \lambda_n(t)\mathrm{d}t\right\}} \qquad \text{利用(5.4)}\\
&=\exp\left\{-\int_A^B \lambda_n(t)\mathrm{d}t\right\}
\end{aligned}
$$

而根据相同的原因, 对应的吸烟者的概率为

$$
\begin{aligned}
&P\{\text{年龄为} A \text{ 的吸烟者能活到年龄} B\}\\
&=\exp\left\{-\int_A^B \lambda_s(t)\mathrm{d}t\right\}=\exp\left\{-2\int_A^B \lambda_n(t)\mathrm{d}t\right\}=\left[\exp\left\{-\int_A^B \lambda_n(t)\mathrm{d}t\right\}\right]^2
\end{aligned}
$$

也就是说, 对于两个年龄相同的人来说, 其中一个吸烟, 另一个不吸烟, 那么吸烟者能存活到一个给定年龄的概率是非吸烟者的相应概率的平方 (而不是一半). 举例来说, 如果 $\lambda_n(t)=1/30, 50\leqslant t\leqslant 60$, 那么一个 50 岁的非吸烟者能活到 60 岁的概率是 $\mathrm{e}^{-1/3}\approx 0.7165$, 而吸烟者的相应概率为 $\mathrm{e}^{-2/3}\approx 0.5134$. ■

5.6 其他连续型分布

5.6.1 Γ 分布

如果随机变量具有密度函数

$$
f(x)=\begin{cases}\dfrac{\lambda \mathrm{e}^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} & x\geqslant 0\\ 0 & x<0\end{cases}
$$

其中 $\Gamma(\alpha)$ 是 Γ 函数, 则称该随机变量具有 Γ 分布, 其参数为 $(\alpha,\lambda), \alpha>0, \lambda>0$. Γ 函数的定义如下:

$$
\Gamma(\alpha)=\int_0^{\infty} \mathrm{e}^{-y} y^{\alpha-1}\mathrm{d}y
$$

分部积分可以得到

$$\Gamma(\alpha) = -\mathrm{e}^{-y}y^{\alpha-1}\big|_0^\infty + \int_0^\infty \mathrm{e}^{-y}(\alpha-1)y^{\alpha-2}\,\mathrm{d}y$$
$$= (\alpha-1)\int_0^\infty \mathrm{e}^{-y}y^{\alpha-2}\mathrm{d}y = (\alpha-1)\Gamma(\alpha-1) \tag{6.1}$$

对于整数 α, 比如说 $\alpha = n$, 可以重复利用公式 (6.1) 得到

$$\Gamma(n) = (n-1)\Gamma(n-1) = (n-1)(n-2)\Gamma(n-2)$$
$$= \cdots = (n-1)(n-2)\cdots 3\cdot 2\Gamma(1)$$

又因 $\Gamma(1) = \int_0^\infty \mathrm{e}^{-x}\mathrm{d}x = 1$, 所以可以得到 $\Gamma(n) = (n-1)!$.

当 α 为一正整数时, 比方说 $\alpha = n$, 参数为 (α, λ) 的 Γ 分布在实践中经常作为某个事件总共要发生 n 次的等待时间的分布而出现. 尤其是, 当 n 个事件是随机发生的, 且满足 4.7 节的三个条件, 那么可以证明要等待总共 n 个事件发生的时间为服从参数为 (n, λ) 的 Γ 分布随机变量. 为了证明这点, 令 T_n 表示第 n 个事件发生的时间, 并注意到 T_n 小于或等于 t 的充要条件是在时刻 t 以前至少发生了 n 个事件, 即时间区间 $[0, t]$ 内发生的事件数 $N(t) \geqslant n$. 因此

$$P\{T_n \leqslant t\} = P\{N(t) \geqslant n\} = \sum_{j=n}^{\infty} P\{N(t) = j\} = \sum_{j=n}^{\infty} \frac{\mathrm{e}^{-\lambda t}(\lambda t)^j}{j!}$$

其中最后一个等式成立是因为在 $[0, t]$ 内发生的事件数服从参数为 λt 的泊松分布, 对上式求导得到 T_n 的密度函数如下

$$f(t) = \sum_{j=n}^{\infty} \frac{\mathrm{e}^{-\lambda t} j(\lambda t)^{j-1}\lambda}{j!} - \sum_{j=n}^{\infty} \frac{\lambda \mathrm{e}^{-\lambda t}(\lambda t)^j}{j!}$$
$$= \sum_{j=n}^{\infty} \frac{\lambda \mathrm{e}^{-\lambda t}(\lambda t)^{j-1}}{(j-1)!} - \sum_{j=n}^{\infty} \frac{\lambda \mathrm{e}^{-\lambda t}(\lambda t)^j}{j!} = \frac{\lambda \mathrm{e}^{-\lambda t}(\lambda t)^{n-1}}{(n-1)!}$$

因此, T_n 服从参数为 (n, λ) 的 Γ 分布. 注意当 $n = 1$ 时, 该分布退化为指数分布.

$\lambda = 1/2, \alpha = n/2$ 的 Γ 分布 (n 为一个正整数) 称为自由度为 n 的 χ^2(读作"卡方") 分布. 在 n 维空间中试图击中某一靶子, 其中各坐标的偏差相互独立且为标准正态分布, 则偏差的平方和服从自由度为 n 的 χ^2 分布. χ^2 分布与正态分布关系密切, 我们将在第 6 章予以详细解释.

例 6a 令 X 为一服从 Γ 分布的随机变量, 其参数为 α 和 λ, 计算

(a) $\mathrm{E}[X]$; (b) $\mathrm{Var}(X)$.

解: (a)

$$E[X] = \frac{1}{\Gamma(\alpha)}\int_0^\infty \lambda x \mathrm{e}^{-\lambda x}(\lambda x)^{\alpha-1}\mathrm{d}x = \frac{1}{\lambda\Gamma(\alpha)}\int_0^\infty \lambda \mathrm{e}^{-\lambda x}(\lambda x)^{\alpha}\mathrm{d}x$$
$$= \frac{\Gamma(\alpha+1)}{\lambda\Gamma(\alpha)} = \frac{\alpha}{\lambda} \qquad \text{利用公式 (6.1)}$$

(b) 先计算 $E[X^2]$, 再利用方差的计算公式可得 $\mathrm{Var}(X)=\frac{\alpha}{\lambda^2}$. 详细证明留作习题. ■

5.6.2 威布尔分布

威布尔分布在工程实践中有着广泛的应用. 最初, 这种分布是在解释疲劳数据时提出的, 但现在它的应用已经扩展到许多其他工程问题中. 特别地, 在有关生存现象的领域中, 它有广泛的应用. 例如, 当某对象适合"最弱链"模型时, 此对象的寿命就服从威布尔分布. 也就是说, 考虑一个有许多部分组成的对象, 并假定当它的任何一部分毁坏时此对象的寿命就终止. 在这样的条件下, 已经 (从理论上和实践上) 证明威布尔分布为这个对象的寿命的分布提供了一个很好的近似.

威布尔分布的分布函数具有如下形式:

$$F(x)=\begin{cases}0 & x\leqslant\nu\\ 1-\exp\left\{-\left(\dfrac{x-\nu}{\alpha}\right)^{\beta}\right\} & x>\nu\end{cases}\tag{6.2}$$

如果一个随机变量的分布函数具有如上形式, 那么称它为具有参数 ν,α 和 β 的威布尔随机变量. 对其求导可得到密度函数为

$$f(x)=\begin{cases}0 & x\leqslant\nu\\ \dfrac{\beta}{\alpha}\left(\dfrac{x-\nu}{\alpha}\right)^{\beta-1}\exp\left\{-\left(\dfrac{x-\nu}{\alpha}\right)^{\beta}\right\} & x>\nu\end{cases}$$

5.6.3 柯西分布

一个随机变量称为服从参数为 $\theta,-\infty<\theta<\infty$ 的柯西分布, 如果其密度函数如下:

$$f(x)=\frac{1}{\pi}\frac{1}{1+(x-\theta)^2}\qquad -\infty<x<\infty$$

例 6b 假设有一束窄窄的光线围绕着某一个中心旋转, 而这个中心位于 y 轴上离坐标原点一个单位处, 当光线停止旋转时, 这束光指向 x 轴上一点 X(若光线并不指向 X 轴, 则重新进行试验). 如图 5.7 所示, X 由 y 轴与光线的夹角 θ 确定. θ 的分布是 $(-\pi/2,\pi/2)$ 上的均匀分布, 这样, X 的分布可由下式进行计算

$$F(x)=P\{X\leqslant x\}=P\{\tan\theta\leqslant x\}=P\{\theta\leqslant\arctan x\}=\frac{1}{2}+\frac{1}{\pi}\arctan x$$

其中最后一个等式是因为 θ 在 $(-\pi/2,\pi/2)$ 上服从均匀分布, 分布函数为

$$P\{\theta\leqslant a\}=\frac{a-(-\pi/2)}{\pi}=\frac{1}{2}+\frac{a}{\pi}\qquad -\frac{\pi}{2}<a<\frac{\pi}{2}$$

因此, X 的密度函数为

$$f(x)=\frac{\mathrm{d}}{\mathrm{d}x}F(x)=\frac{1}{\pi(1+x^2)}\qquad -\infty<x<\infty$$

这样, 我们可以看到 X 服从柯西分布[①]. ■

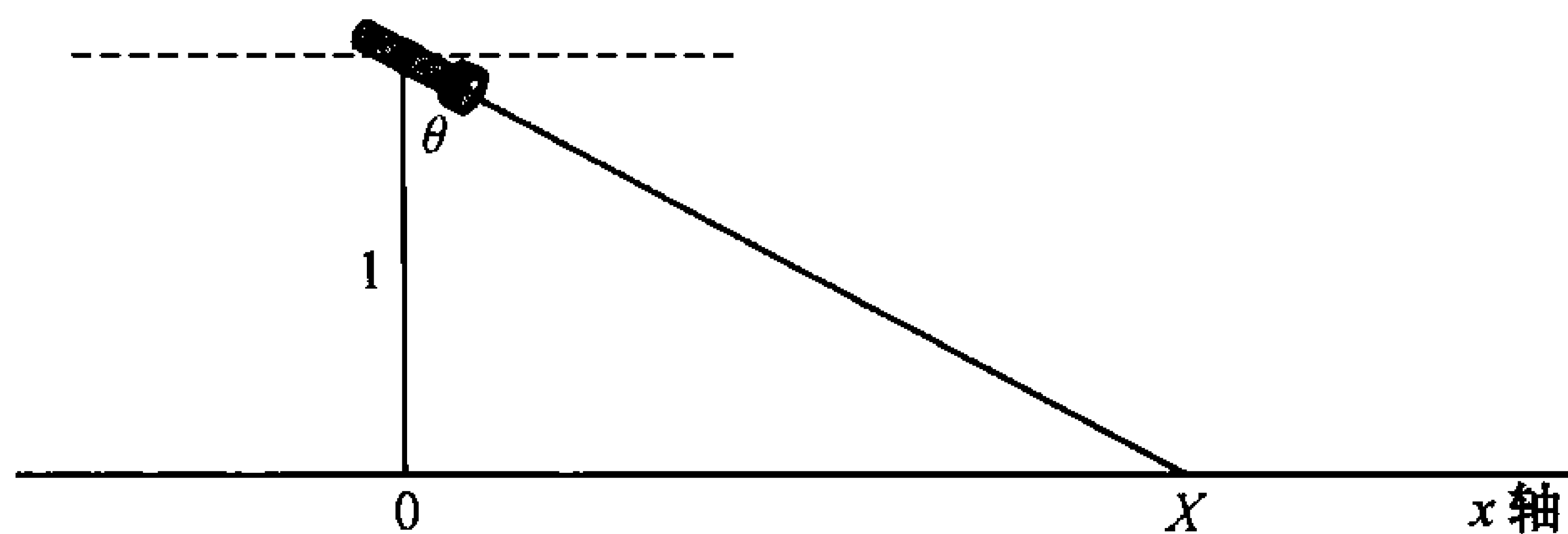


图 5.7

5.6.4 β 分布

如果一个随机变量的密度函数如下, 那么称它服从 β 分布:

$$f(x)=\begin{cases}\dfrac{1}{B(a,b)}x^{a-1}(1-x)^{b-1} & 0<x<1\\ 0 & \text{其他}\end{cases}$$

其中

$$B(a,b)=\int_0^1 x^{a-1}(1-x)^{b-1}\,\mathrm{d}x$$

β 分布通常用来为取值于某有限区间 $[c,d]$ 的随机现象建立模型. 当然, 如果令 c 为原点, 而 $d-c$ 为度量单位, 那么可将取值区间转化为 $[0,1]$.

当 $a=b$ 时, β 分布的密度函数关于 $x=1/2$ 对称, 随着公共值 a 的增大, 取值于 $1/2$ 附近的权重会越来越大 (见图 5.8). 当 $b>a$ 时, 密度函数向左偏斜 (也即取小值的可能性更大); 当 $a>b$ 时, 密度函数向右偏斜 (见图 5.9).

$B(a,b)=\int_0^1 x^{a-1}(1-x)^{b-1}\,\mathrm{d}x$ 称为 β 函数. 可以证明, β 函数与 Γ 函数之间存在以下关系:

$$B(a,b)=\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \tag{6.3}$$

利用 (6.1) 式和 (6.3) 式, 易证, 如果 X 为一参数为 a 和 b 的 β 随机变量, 那么

$$E[X]=\frac{a}{a+b}\qquad \mathrm{Var}(X)=\frac{ab}{(a+b)^2(a+b+1)}$$

注释 (6.3) 式的证明将会在第 6 章例 7c 中给出.

---

① 可用下面的方法证明等式 $\dfrac{\mathrm{d}}{\mathrm{d}x}(\arctan x)=1/(1+x^2)$: 令 $y=\arctan x$, 那么 $\tan y=x$, 因此:

$$1=\frac{\mathrm{d}}{\mathrm{d}x}(\tan y)=\frac{\mathrm{d}}{\mathrm{d}y}(\tan y)\frac{\mathrm{d}y}{\mathrm{d}x}=\frac{\mathrm{d}}{\mathrm{d}y}\left(\frac{\sin y}{\cos y}\right)\frac{\mathrm{d}y}{\mathrm{d}x}=\left(\frac{\cos^2 y+\sin^2 y}{\cos^2 y}\right)\frac{\mathrm{d}y}{\mathrm{d}x}$$

或者

$$\frac{\mathrm{d}y}{\mathrm{d}x}=\frac{\cos^2 y}{\sin^2 y+\cos^2 y}=\frac{1}{\tan^2 y+1}=\frac{1}{x^2+1}$$

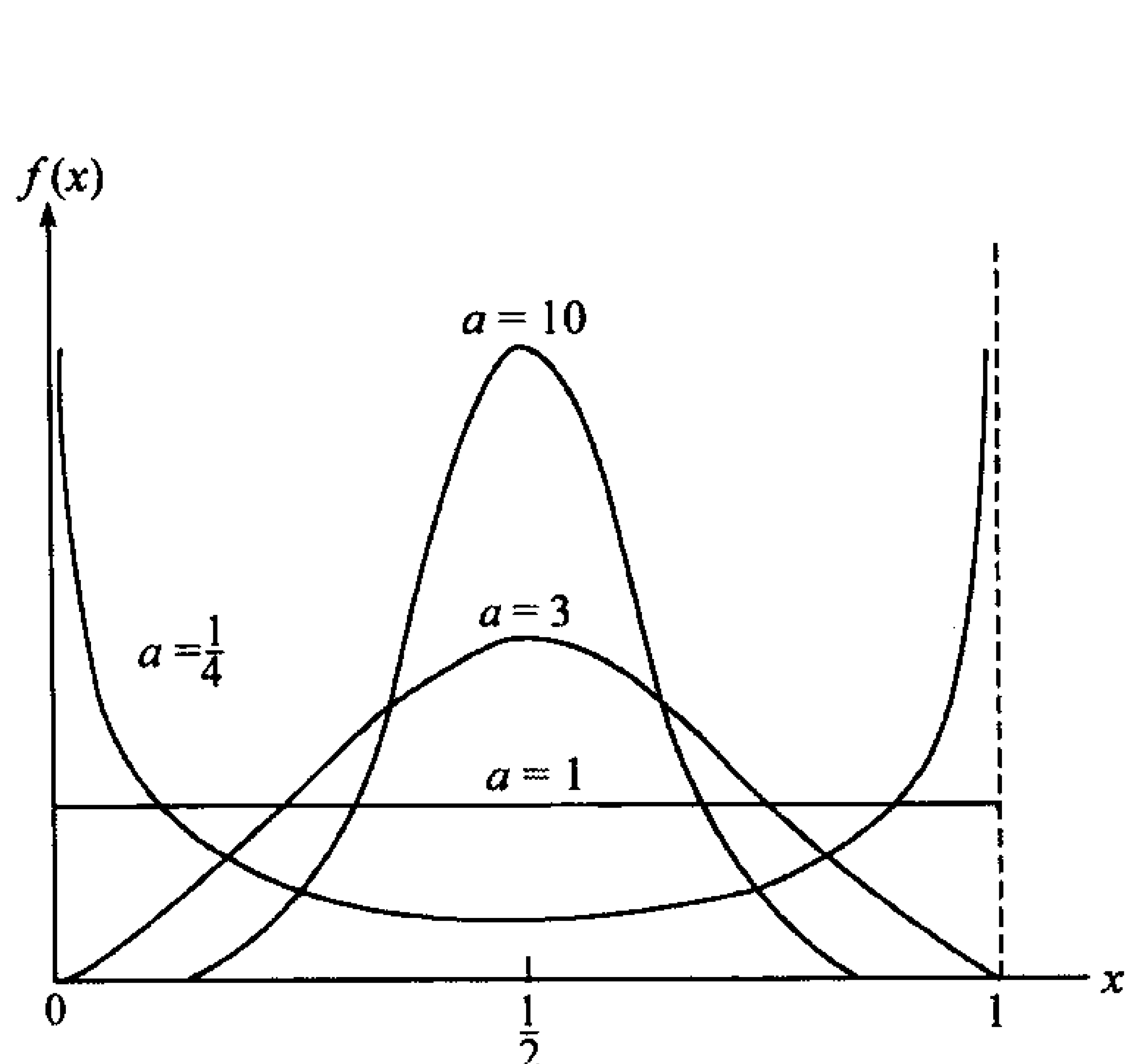


图 5.8 参数为 (a,b) 的 β 分布，其中 $a=b$

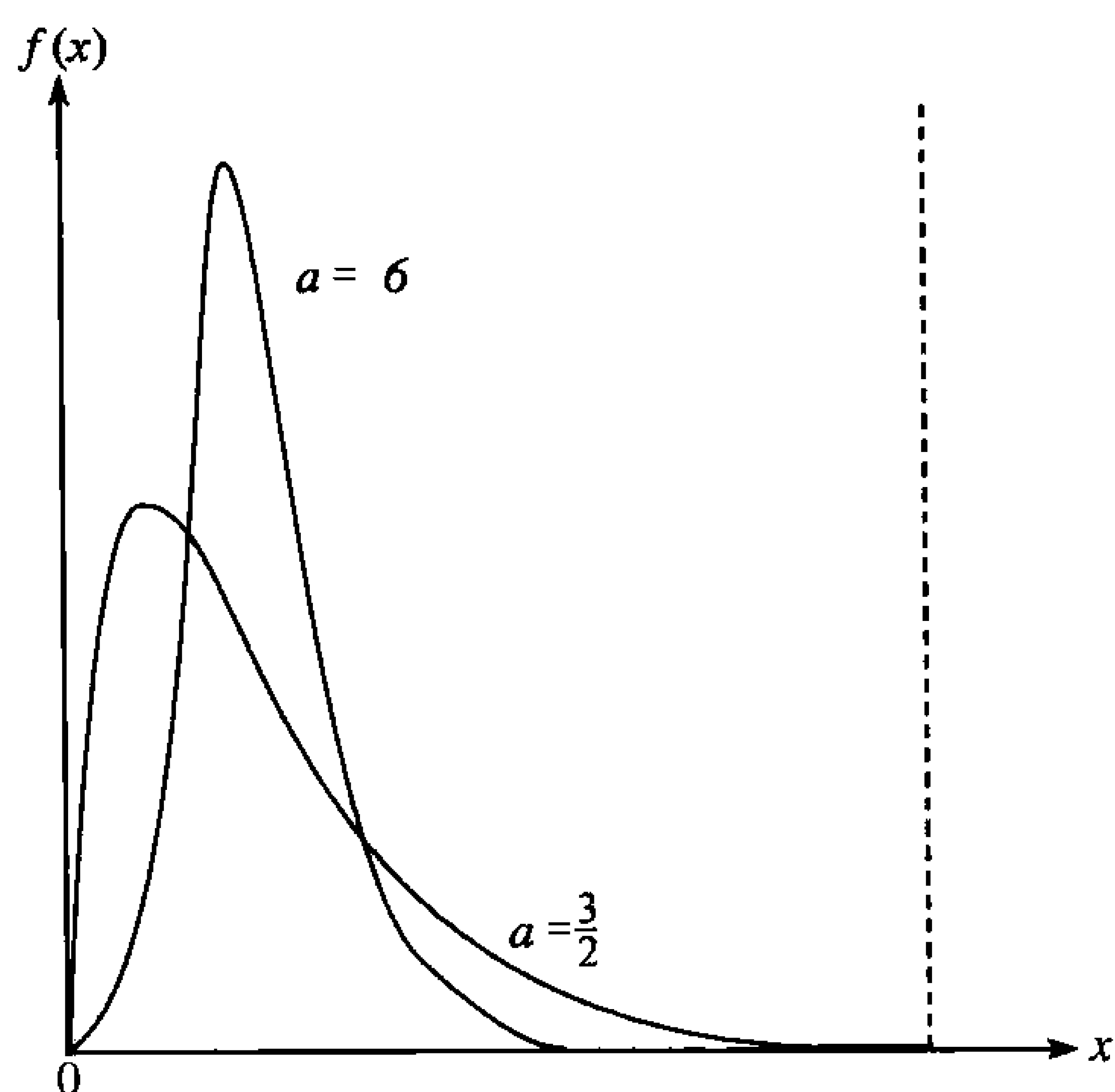


图 5.9 参数为 (a,b) 的 β 分布，其中 $a/(a+b)=1/20$

5.7 随机变量函数的分布

当知道一个随机变量的概率分布后，经常会感兴趣于求它的一些函数的分布. 比如，假设我们已经知道了 X 的分布，需要计算 $g(X)$ 的分布. 要做到这一点，需要将事件 $g(X)\leqslant y$ 表达为 X 属于某集合的形式，我们将通过一些例子阐述这点.

例 7a 令 X 为 $(0,1)$ 上服从均匀分布的随机变量. 我们可以如下得到随机变量 $Y=X^n$ 的分布. 对于 $0\leqslant y\leqslant 1$，有

$$F_Y(y)=P\{Y\leqslant y\}=P\{X^n\leqslant y\}=P\{X\leqslant y^{1/n}\}=F_X(y^{1/n})=y^{1/n}$$

这样，Y 的密度函数为

$$f_Y(y)=\begin{cases}\dfrac{1}{n}y^{1/n-1} & 0\leqslant y\leqslant 1\\ 0 & \text{其他}\end{cases}$$ ■

例 7b 如果 X 是一个连续型随机变量，其密度函数为 f_X，那么 $Y=X^2$ 的分布可以如下得到，对于 $y\geqslant 0$，有

$$F_Y(y)=P\{Y\leqslant y\}=P\{X^2\leqslant y\}=P\{-\sqrt{y}\leqslant X\leqslant\sqrt{y}\}=F_X(\sqrt{y})-F_X(-\sqrt{y})$$

求导可得：$f_Y(y)=\dfrac{1}{2\sqrt{y}}[f_X(\sqrt{y})+f_X(-\sqrt{y})], y\geqslant 0$. ■

例 7c 如果 X 具有密度函数 f_X，那么 $Y=|X|$ 的密度函数可以如下得到，对于 $y\geqslant 0$，有

$$F_Y(y)=P\{Y\leqslant y\}=P\{|X|\leqslant y\}=P\{-y\leqslant X\leqslant y\}=F_X(y)-F_X(-y)$$

因此, 对其求导可得: $f_Y(y) = f_X(y) + f_X(-y), y \geqslant 0$. ■

以上例 7a 到例 7c 中使用的方法可用来证明定理 7.1.

定理 7.1 令 X 为一连续型随机变量, 其密度函数为 f_X. 假设 $g(x)$ 为一严格单调 (递增或递减) 且可微 (因此必连续) 函数, 那么随机变量 $Y = g(X)$ 的密度函数为

$$f_Y(y) = \begin{cases} f_X[g^{-1}(y)]\left|\dfrac{\mathrm{d}}{\mathrm{d}y}g^{-1}(y)\right| & \text{如果存在某}x\text{, 使得}y = g(x) \\ 0 & \text{如果对一切}x, y \neq g(x) \end{cases}$$

其中 $g^{-1}(y)$ 定义为满足 $g(x) = y$ 的 x 值.

我们在 $g(x)$ 为递增函数的情形下证明定理 7.1.

证明: 假设对某些 x 来说, 有 $y = g(x)$. 那么, 令 $Y = g(X)$,

$$F_Y(y) = P\{g(X) \leqslant y\} = P\{X \leqslant g^{-1}(y)\} = F_X(g^{-1}(y))$$

求导可得

$$f_Y(y) = f_X(g^{-1}(y))\frac{\mathrm{d}}{\mathrm{d}y}g^{-1}(y)$$

由于 $g^{-1}(y)$ 非降, 因此导数非负.

若 $y \neq g(x)$ 对任意 x 成立, 那么 $F_Y(y)$ 等于 0 或 1, 无论 $F_Y(y) = 0$ 还是 $F_Y(y) = 1$, 均有 $f_Y(y) = 0$. □

例 7d 设 X 为一非负连续随机变量, 其密度函数为 f, 令 $Y = X^n$, 计算 Y 的密度函数 $f_Y(y)$.

解: 令 $g(x) = x^n$, 那么

$$g^{-1}(y) = y^{1/n}$$

且

$$\frac{\mathrm{d}}{\mathrm{d}y}\{g^{-1}(y)\} = \frac{1}{n}y^{1/n-1}$$

利用定理 7.1, 可得

$$f_Y(y) = \frac{1}{n}y^{1/n-1}f(y^{1/n})$$

当 $n = 2$ 时,

$$f_Y(y) = \frac{1}{2\sqrt{y}}f(\sqrt{y})$$

这与例 7b 的结论是一致的 (因 $X \geqslant 0$). ■

小 结

一个随机变量 X 称为连续型的, 如果存在一个 x 的非负函数 f(称为密度函数), 满足: 对于任一集合 B, 有

$$P\{X \in B\} = \int_B f(x)\mathrm{d}x$$

如果 X 是连续型的, 那么其分布函数 F 在 $f(x)$ 的连续点处可导, 且

$$\frac{\mathrm{d}}{\mathrm{d}x}F(x) = f(x)$$

连续型随机变量 X 的期望值定义如下:

$$E[X] = \int_{-\infty}^{\infty} xf(x)\mathrm{d}x$$

对于任一函数 g, 有一个有用的恒等式:

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)\mathrm{d}x$$

正如离散型情形一样, 随机变量 X 的方差定义为如下:

$$\mathrm{Var}(X) = E[(X - E[X])^2]$$

随机变量 X 称为服从区间 (a,b) 上的**均匀分布**, 如果其密度函数为

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a \leqslant x \leqslant b \\ 0 & \text{其他} \end{cases}$$

其期望和方差分别是

$$E[X] = \frac{a+b}{2} \qquad \mathrm{Var}(X) = \frac{(b-a)^2}{12}$$

随机变量 X 称为服从参数为 μ 和 σ^2 的**正态随机变量**, 如果其密度函数为

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-(x-\mu)^2/2\sigma^2} \qquad -\infty < x < \infty$$

可以证明

$$\mu = E[X] \qquad \sigma^2 = \mathrm{Var}(X)$$

如果 X 服从均值为 μ、方差为 σ^2 的正态分布, 那么如下定义的 Z:

$$Z = \frac{X-\mu}{\sigma}$$

也是服从正态分布的随机变量, 其均值为 0, 方差为 1. 这样的随机变量称为**标准正态随机变量**. 有关 X 的概率可以通过标准正态随机变量 Z 进行计算, 而 Z 的分布函数可以通过查表 5.1 或从网站上得到.

参数为 $(n,\ p)$ 的二项分布, 当 n 足够大时, 可以近似为均值为 np, 方差为 $np(1-p)$ 的正态分布.

一个随机变量称为参数为 λ 的**指数随机变量**, 如果其密度函数为如下形式:

$$f(x)=\begin{cases}\lambda e^{-\lambda x} & x\geqslant 0\\ 0 & \text{其他}\end{cases}$$

其期望值和方差分别为:

$$E[X]=\frac{1}{\lambda} \qquad \mathrm{Var}(X)=\frac{1}{\lambda^2}$$

一个只有指数随机变量才具有的重要性质是**无记忆性**, 也即对于正数 s 和 t, 有

$$P\{X>s+t|X>t\}=P\{X>s\}$$

如果 X 表示某个零件的寿命, 那么无记忆性就说明了对任意 t, 年龄为 t 的零件的剩余寿命同一个新的零件的寿命的分布是一样的.

令 X 为一个非负连续型随机变量, 其分布函数为 F, 密度函数为 f, 那么函数

$$\lambda(t)=\frac{f(t)}{1-F(t)} \qquad t\geqslant 0$$

称为 F 的**危险率函数**, 或者**失效率函数**. 如果我们认为 X 是某个零件的寿命, 那么对于一个很小的值 $\mathrm{d}t$, $\lambda(t)\,\mathrm{d}t$ 近似为年龄为 t 的零件在 $\mathrm{d}t$ 时间内会失效的概率. 如果 F 是参数为 λ 的指数分布, 那么

$$\lambda(t)=\lambda \qquad t\geqslant 0$$

另外, 指数分布是唯一的失效率为常数的分布.

一个随机变量称为参数为 (α, λ) 的 Γ 随机变量, 如果其密度函数等于

$$f(x)=\frac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} \qquad x\geqslant 0$$

$\Gamma(\alpha)$ 称为 Γ 函数, 定义为

$$\Gamma(\alpha)=\int_0^{\infty} e^{-x}x^{\alpha-1}\mathrm{d}x$$

Γ 随机变量的期望和方差分别如下:

$$E[X]=\frac{\alpha}{\lambda} \qquad \mathrm{Var}(X)=\frac{\alpha}{\lambda^2}$$

随机变量称为服从参数为 (a,b) 的β 分布, 如果其密度函数为

$$f(x)=\frac{1}{B(a,b)}x^{a-1}(1-x)^{b-1} \qquad 0\leqslant x\leqslant 1$$

常数 $B(a,b)$ 的定义如下:

$$B(a,b)=\int_0^1 x^{a-1}(1-x)^{b-1}\mathrm{d}x$$

β 随机变量的期望和方差分别为

$$E[X]=\frac{a}{a+b} \qquad \mathrm{Var}(X)=\frac{ab}{(a+b)^2(a+b+1)}$$

习 题

1. 令 X 是一随机变量, 其密度函数为

$$f(x)=\begin{cases}C(1-x^2) & -1<x<1\\ 0 & 其他\end{cases}$$

(a) C 的值是多少? (b) 求 X 的分布函数.

2. 一个系统由一个元件和它的替换元件组成, 若这个元件的寿命为随机变量 X, X 的密度函数由下式给出 (单位为月):

$$f(x)=\begin{cases}Cx\mathrm{e}^{-x/2} & x>0\\ 0 & x\leqslant 0\end{cases}$$

问这个系统能维持 5 个月的概率有多大?

3. 考虑函数

$$f(x)=\begin{cases}C(2x-x^3) & 0<x<\frac{5}{2}\\ 0 & 其他\end{cases}$$

f 能是一个概率密度函数么? 如果是, 求 C. 如果 $f(x)$ 为如下的函数呢?

$$f(x)=\begin{cases}C(2x-x^2) & 0<x<\frac{5}{2}\\ 0 & 其他\end{cases}$$

4. 设随机变量 X 表示某个电子设备的寿命 (单位: 小时), 其密度函数如下:

$$f(x)=\begin{cases}\dfrac{10}{x^2} & x>10\\ 0 & x\leqslant 10\end{cases}$$

(a) 求 $P\{X>20\}$; (b) X 的分布函数是什么?

(c) 6 个类似的设备中, 至少有三个寿命超过 15 小时的概率是多大? 其中作了什么假设?

5. 一个加油站每周补给一次油. 如果它每周的销售量 (单位: 千加仑) 为一随机变量, 其密度函数为

$$f(x)=\begin{cases}5(1-x)^4 & 0<x<1\\ 0 & 其他\end{cases}$$

试问油罐需要多大, 才能把一周内断油的概率控制为 0.01?

6. 如果 X 的密度函数如下, 求 $E[X]$:

(a) $f(x) = \begin{cases} \frac{1}{4}x\mathrm{e}^{-x/2} & x > 0 \\ 0 & \text{其他} \end{cases}$ (b) $f(x) = \begin{cases} c(1-x^2) & -1 < x < 1 \\ 0 & \text{其他} \end{cases}$

(c) $f(x) = \begin{cases} \dfrac{5}{x^2} & x > 5 \\ 0 & x \leqslant 5 \end{cases}$

7. X 的概率密度函数如下:

$$f(x) = \begin{cases} a + bx^2 & 0 \leqslant x \leqslant 1 \\ 0 & \text{其他} \end{cases}$$

如果 $E[X] = \frac{2}{3}$, 求 a 和 b.

8. 某种电子管的使用寿命 (单位: 小时) 是一个随机变量, 其密度函数为

$$f(x) = x\mathrm{e}^{-x} \qquad x \geqslant 0$$

求电子管的寿命的期望值.

9. 考虑第 4 章的例 4b, 但现在假设季度需求量为一连续型随机变量, 其密度函数为 f, 指出最优储存量 s^* 应满足下面的条件:

$$F(s^*) = \frac{b}{b+\ell}$$

其中 b 是每个单位销售量的利润, ℓ 是每一个未销售单位的损失, F 是每季需求量的分布函数.

10. 从早上 7 点开始, 每隔 15 分钟都有一趟列车开往 A 地, 而从 7:05 开始, 每隔 15 分钟有一趟列车开往 B 地.

(a) 如果某位乘客到达车站的时间服从在 7 点到 8 点之间的均匀分布. 他到达车站以后, 无论下一趟进站的是开往 A 地还是开往 B 的, 他立刻上那一列车. 问他乘上开往 A 地的列车的概率是多大?

(b) 如果该乘客达到达车站的时间服从 7:10 到 8:10 之间的均匀分布呢?

11. 从长为 L 的线段上随机选一点, 计算短的那一截相对于长的那一截的比例小于 1/4 的概率.

12. 一辆长途汽车运行在相距 100 英里的 A 和 B 两城市之间. 如果汽车抛锚的话, 抛锚地点距 A 的距离应该服从 (0, 100) 的均匀分布. 又已知共有三个汽车服务站, 分别设在城市 A、城市 B 以及两城之间的中点. 有人建议这三个服务站应该分别设在离城市 A 距离 25, 50, 75 英里处, 这样的设置效率更高, 你同意吗? 为什么?

13. 你于 10 点到达公共汽车站, 并且已知汽车到达的时间在 10 点和 10 点半之间均匀分布.

(a) 你等待时间超过 10 分钟的概率是多大?

(b) 如果 10:15 时, 汽车还没有来, 那么你至少还要等待 10 分钟的概率是多大?

14. 令 X 为 (0, 1) 上均匀随机变量, 利用命题 2.1 计算 $E[X^n]$, 并利用期望的定义验证该结论.

15. 如果 X 为服从参数为 $\mu = 10$ 和 $\sigma^2 = 36$ 的正态随机变量, 求

(a) $P\{X > 5\}$; (b) $P\{4 < X < 16\}$; (c) $P\{X < 8\}$;

(d) $P\{X < 20\}$; (e) $P\{X > 16\}$.

16. 某个地区的年降雨量 (单位: 英寸) 服从参数为 $\mu = 40$ 和 $\sigma = 4$ 的正态分布. 那么从今年开始, 10 年以内, 每年的降雨量不超过 50 英寸的概率有多大? 你作了什么假设?

17. 某人向某个目标射击, 如果击中目标 1 英寸范围内, 那么获得 10 分; 如果击中目标 1 英寸到 3 英寸之间, 获得 5 分; 如果果击中目标 3 英寸到 5 英寸之间, 获得 3 分. 如果击中点与靶心的距离服从 (0, 10) 上的均匀分布, 求获得分数的期望值.

18. 假设 X 是一个正态随机变量, 其均值为 5. 如果 $P\{X > 9\} = 0.2$, 求方差的近似值.

19. 令 X 为一正态随机变量, 均值为 12, 方差为 4. 求满足条件 $P\{X > c\} = 0.10$ 的 c 值.

20. 在某社区内, 有 65% 的人支持提高教育税. 如果随机抽取 100 人, 估算以下事件概率:

(a) 至少 50 人支持该建议; (b) 支持该建议的人在 60 人到 70 人 (含) 之间;

(c) 少于 75 人支持该建议.

21. 假设 25 岁男人的身高 (单位: 英寸) 是参数为 $\mu = 71$, $\sigma^2 = 6.25$ 的正态随机变量. 那么身高超过 6 英尺 2 英寸的比例有多大? 在身高皆超过 6 英尺的成人俱乐部里, 身高超过 6 英尺 5 英寸的比例有多大?

22. 有一种锻造的铜铝合金零件, 其上有一开口. 开口宽度具有正态分布, 期望 $\mu = 0.9$(英寸), $\sigma = 0.003$. 产品的工艺标准为 0.9000 ± 0.005.

(a) 零件的废品率为多少?

(b) 若容许的废品率为 0.01, σ 应控制在多大才能满足这个要求?

23. 掷一枚均匀骰子 1000 次, 求点数 6 出现的次数在 150 到 200 次 (含) 之间的概率的近似值, 如果点数 6 正好出现了 200 次, 求点数 5 出现的次数小于 150 次的概率.

24. 某半导体工厂生产的计算机芯片的寿命服从正态分布, 参数为 $\mu = 1.4 \times 10^6$ 小时, $\sigma = 3 \times 10^5$ 小时. 求一批 100 个芯片内, 包含至少 20 个芯片, 其寿命都小于 1.8×10^6 的概率的近似值.

25. 某工厂生产的每个元件都以 0.95 的概率被接收, 且各元件是否被接收是相互独立的 (即不是成批接收). 问下一批 150 个元件内最多 10 件不被接受的概率的近似值.

26. 某工厂生产两种硬币, 一种是均匀硬币, 另一种在抛掷时正面朝上的概率为 55%. 现有一枚该工厂的硬币, 但是不知道是哪一种. 为了确认是哪一种, 我们进行如下统计检验: 掷这枚硬币 1000 次, 如果正面朝上的次数超过 525 次 (含), 那么我们认为该硬币不是均匀的, 否则, 就认为它是均匀的. 如果该硬币确实是均匀的, 我们将会得到错误结论的概率是多大? 如果该硬币不是均匀的呢?

27. 在 10 000 次独立地掷硬币过程中, 如果正面朝上的次数超过 5800 次, 那么有理由认为这枚硬币不是均匀的. 试解释之.

28. 人群中有 12% 的人为左撇子, 估算在一个有 200 人的学校中, 左撇子人数超过 20 的概率的近似值. 在这个问题中, 你作了什么样的假设?

29. 某股价波动模型认为, 如果目前的股价为 s, 那么一个周期后, 股价变成 us 的概率为 p, 变成 ds 的概率为 $1 - p$. 假设各周期的股价波动是相互独立的, 试估算未来 1000 个周期后, 股价上涨 30% 的概率, 其中 $u = 1.012, d = 0.990, p = 0.52$.

30. 一个存储部件分为两个区域, 白区和黑区. 当随机地从白区读取数据, 所得到的数据长度可用一个正态随机变量表示, 其期望 $\mu = 4$, 方差为 $\sigma^2 = 4$. 当随机地从黑区读取数据, 所得到的数据长度的参数为 $\mu = 6, \sigma^2 = 9$. 现在随机地从存储器读取一个数据, 其长度为 5. 现设黑区所占的比例为 α. 当我们取得数据长度为 5 时, 说数据来自白区, 或黑区, 都会有一定的犯错误概率. 当 α 多大时, 两种犯错误的概率一样大?

31. (a) 某救火站设在一段长为 $A, A < \infty$ 的路上. 如果着火点均匀分布在区间 $(0, A)$ 上, 那么救火站应该设在何处, 使得到着火点的期望距离最小. 也即, 计算 a, 使得 $E[|X - a|]$ 达到最小值, 其中 $x \in (0, A)$ 为着火点.

(b) 现假设该段路的长度为无限长 ——从原点 0 一直延伸到 ∞, 如果着火点距离原点的距离 X 服从参数为 λ 的指数分布, 那么救火站应该设在何处, 如果我们要使得 $E[|X - a|]$ 极小化?

32. 修理某机器所需的时间 (单位: 小时) 是参数为 $\lambda = 1/2$ 的指数随机变量, 试问

(a) 修理时间超过 2 小时的概率是多大?

(b) 若已持续修理了 9 小时, 总共需要 10 小时才能修好的概率是多大?

33. 收音机工作时间 (单位: 年) 服从参数为 $\lambda = 1/8$ 的指数分布, 琼斯买了一台收音机, 问 8 年以后它仍能工作的概率是多大?

34. 琼斯估计, 一辆汽车在报废之前能行驶的里程 (单位: 千英里) 是以 $\lambda = 1/20$ 为参数的指数随机变量, 史密斯有一辆自称是只行驶过 10 000 英里的旧车, 如果琼斯买下这辆汽车, 按她的上述估计, 至少还能行驶 20 000 英里的概率是多大? 若将指数分布这一假设改为 (0, 40) 上的均匀分布, 上述概率又是多大?

35. 对于一个年龄为 t 的男性吸烟者来说, 患肺癌的危险率函数 $\lambda(t)$ 为

$$\lambda(t) = 0.027 + 0.000\,25(t - 40)^2 \qquad t \geqslant 40$$

假定一个 40 岁的男性吸烟者没有患肺癌, 那么他活到如下岁数仍不患肺癌的概率是多大?

(a) 50 岁. (b) 60 岁.

36. 假设某个元件的寿命分布的失效率函数为 $\lambda(t) = t^3, t > 0$. 求以下事件概率:

(a) 寿命将超过 2 年; (b) 寿命在 0.4 年到 1.4 年之间;

(c) 某个已经用了 1 年的元件, 还能使用 1 年.

37. 如果 X 服从 $(-1, 1)$ 上均匀分布, 求

(a) $P\{|X| > \frac{1}{2}\}$; (b) 随机变量 $|X|$ 的密度函数.

38. 如果 Y 服从 $(0, 5)$ 上的均匀分布, 那么方程 $4x^2 + 4xY + Y + 2 = 0$ 的两个根都为实数的概率有多大?

39. 如果 X 是一个指数随机变量, 其参数为 $\lambda = 1$, 求随机变量 Y 的密度函数, 其中 $Y = \ln X$.

40. 如果 X 服从 $(0, 1)$ 上的均匀分布, 求 $Y = \mathrm{e}^X$ 的密度函数.

41. 计算 $R = A \sin\theta$ 的分布函数, 其中 A 为一给定常数, θ 服从 $(-\pi/2, \pi/2)$ 上均匀分布. 这样的随机变量 R 出现在弹道学理论中. 如果一枚炮弹从原点开始发射, 仰角为 α, 速度为 v, 则命中点 R 可以表示为 $R = (v^2/g) \sin 2\alpha$, 其中 g 为地球引力常数, 等于 9.8 米/秒 ^{2}.

理论习题

1. 平衡状态的气体分子的速度是一个随机变量, 其密度函数为

$$f(x)=\begin{cases}ax^2\mathrm{e}^{-bx^2} & x\geqslant 0\\ 0 & x<0\end{cases}$$

其中 $b=m/2kT$, k,T 和 m 分别表示波尔兹曼常数, 绝对温度以及分子的质量. 求常数 a 的值.

2. 证明:

$$E[Y]=\int_0^\infty P\{Y>y\}\mathrm{d}y-\int_0^\infty P\{Y<-y\}\mathrm{d}y$$

提示: 证明

$$\int_0^\infty P\{Y<-y\}\mathrm{d}y=-\int_{-\infty}^0 xf_Y(x)\mathrm{d}x$$
$$\int_0^\infty P\{Y>y\}\mathrm{d}y=\int_0^\infty xf_Y(x)\mathrm{d}x$$

3. 如果 X 的密度函数为 f, 证明

$$E[g(X)]=\int_{-\infty}^\infty g(x)f(x)\mathrm{d}x$$

提示: 利用理论习题 2, 先从等式

$$E[g(x)]=\int_0^\infty P\{g(X)>y\}\mathrm{d}y-\int_0^\infty P\{g(X)<-y\}\mathrm{d}y$$

开始, 然后利用本书中当 $g(X)\geqslant 0$ 的情况下的证明方法继续下去.

4. 证明推论 2.1.

5. 对于一个非负随机变量 Y, 有

$$E[Y]=\int_0^\infty P\{Y>t\}\mathrm{d}t$$

利用此结论来证明: 对于非负随机变量 X, 有

$$E[X^n]=\int_0^\infty nx^{n-1}P\{X>x\}\mathrm{d}x$$

提示: 从

$$E[X^n]=\int_0^\infty P\{X^n>t\}\mathrm{d}t$$

开始, 然后作变量替换 $t=x^n$.

6. 定义一系列事件 $E_a, 0<a<1$, 满足: 对任意 a 有 $P(E_a)=1$, 但 $P(\bigcap_a E_a)=0$.

提示: 令 X 为 $(0,1)$ 上均匀分布, 利用 X 定义每个 E_a.

7. 称

$$\mathrm{SD}(X)=\sqrt{\mathrm{Var}(X)}$$

为 X 的标准差. 若 X 的方差为 σ^2, 求 $\mathrm{SD}(aX+b)$.

8. 令 X 为取值为 0 到 c 之间的随机变量, 也即 $P\{0\leqslant X\leqslant c\}=1$, 证明:

$$\mathrm{Var}(X)\leqslant\frac{c^2}{4}$$

提示: 先证明

$$E[X^2]\leqslant cE[X]$$

然后利用它证明

$$\mathrm{Var}(X)\leqslant c^2[\alpha(1-\alpha)] \qquad 其中\alpha=\frac{E[X]}{c}$$

9. 设 Z 为标准正态随机变量, 证明: 对于 $x>0$, 有

(a) $P\{Z>x\}=P\{Z<-x\}$; (b) $P\{|Z|>x\}=2P\{Z>x\}$;

(c) $P\{|Z|<x\}=2P\{Z<x\}-1$.

10. 令 $f(x)$ 表示参数为 μ 和 σ^2 的正态随机变量的密度函数, 证明 $\mu-\sigma$ 和 $\mu+\sigma$ 为该函数的拐点, 也即证明: 当 $x=\mu-\sigma$ 或 $x=\mu+\sigma$ 时 $f''(x)=0$.

11. 令 Z 为标准正态随机变量, g 为可微函数，其导函数为 g'.

(a) 证明 $E[g'(Z)]=E[Zg(Z)]$

(b) 证明 $E[Z^{n+1}]=nE[Z^{n-1}]$

(c) 求 $E[Z^4]$

12. 设 X 是参数为 λ 的指数随机变量, 利用理论习题 5 中的等式，求出 $E[X^2]$.

13. 设某连续随机变量的分布函数为 F. 则满足 $F(m)=1/2$ 的 m 称为这个随机变量的中位数. 也即, 随机变量取值大于中位数的概率与取值小于中位数的概率是一样的. 当随机变量 X 服从以下分布时, 求其中位数:

(a) 在 (a,b) 区间上均匀分布; (b) 参数为 μ 和 σ^2 的正态分布;

(c) 参数为 λ 的指数分布.

14. 设随机变量的分布密度为 f, 使得 $f(x)$ 达到最大值的点称为它的众数. 求以上理论习题 13 中的 (a), (b), (c) 里的随机变量的众数;

15. 如果 X 为参数为 λ 的指数随机变量, 对于 $c>0$, 证明 cX 服从参数为 λ/c 的指数分布.

16. 当 X 服从 $(0,a)$ 上均匀分布时, 求其危险率函数.

17. 如果 X 的危险率函数为 $\lambda_X(t)$, 计算 aX 的危险率函数, 其中 a 为一正常数.

18. 证明 Γ 分布的密度函数积分值为 1.

19. 如果 X 为均值为 $1/\lambda$ 的指数随机变量, 证明

$$E[X^k]=\frac{k!}{\lambda^k} \qquad k=1,2,\cdots$$

提示: 利用 Γ 分布密度函数来计算.

20. 证明:

$$\mathrm{Var}(X)=\frac{\alpha}{\lambda^2}$$

其中 X 为参数为 (α,λ) 的 Γ 分布随机变量.

21. 证明 $\Gamma(\frac{1}{2})=\sqrt{\pi}$.

提示: $\Gamma(\frac{1}{2})=\int_0^{\infty}\mathrm{e}^{-x}x^{-1/2}\mathrm{d}x$, 做变量替换 $y=\sqrt{2x}$, 然后将结果与正态分布联系起来.

22. 求参数为 (α,λ) 的 Γ 分布随机变量的危险率函数, 并证明当 $\alpha\geqslant 1$ 时危险率函数递增, 而 $\alpha\leqslant 1$ 时递减.

23. 求威布尔随机变量的危险率函数, 并证明当 $\beta\geqslant 1$ 时递增, 而 $\beta\leqslant 1$ 时递减.

24. 如果 $F(\cdot)$ 为威布尔分布函数, 指出 $\ln(\ln(1-F(x))^{-1})$ 对于 $\ln x$ 是一条斜率为 β 的直线. 且证明, 当 $\nu=0$ 时, 大约 63.2% 的观测值都将小于 α.

25. 令

$$Y=\left(\frac{X-\nu}{\alpha}\right)^{\beta}$$

证明: 如果 X 是服从参数为 ν,α,β 的威布尔分布的随机变量, 那么 Y 就是服从参数为 $\lambda=1$ 的指数随机变量, 反之也成立.

26. 如果 X 是服从参数为 (a,b) 的 β 分布的随机变量, 证明:

$$E[X]=\frac{a}{a+b}\qquad \mathrm{Var}(X)=\frac{ab}{(a+b)^2(a+b+1)}$$

27. 如果 X 服从 (a,b) 上的均匀分布, 那么和 X 有线性关系, 且服从 $(0,1)$ 上均匀分布的随机变量是什么?

28. 考虑参数为 (a,b) 的 β 分布, 证明

(a) 当 $\alpha>1$ 和 $b>1$ 时, 密度函数为单众数 (即密度函数只有一个众数), 其众数为 $(a-1)/(a+b-2)$;

(b) 如果 $a\leqslant 1,b\leqslant 1$, 且 $a+b<2$, 那么密度函数或者是单众数的, 此时众数为 0 或 1, 或者为 U 型函数, 0 和 1 都是众数;

(c) 当 $a=1=b$ 时, $[0,1]$ 上所有点都是众数.

29. 设 X 为连续型随机变量, 其分布函数为 F. 定义随机变量 Y: $Y=F(X)$. 证明: Y 服从 $(0,1)$ 上的均匀分布.

30. 设 X 的密度函数为 f_X, 求随机变量 Y 的概率密度函数, 其中 $Y=aX+b$.

31. 设 X 为服从参数为 μ 和 σ^2 的正态随机变量, 求 $Y=\mathrm{e}^X$ 的密度函数. 随机变量 Y 称为服从参数为 μ 和 σ^2 的对数正态 (因为 $\ln Y$ 服从正态分布) 随机变量.

32. 令 X 和 Y 为相互独立的随机变量, 它们都在 $1,2,\cdots,10^N$ 上等可能取值, 其中 N 是一个足够大的数. 令 D 表示 X 和 Y 的最大公约数, 且令 $Q_k=P\{D=k\}$.

(a) 用直观的方法论证 $Q_k=\dfrac{1}{k^2}Q_1$;

提示: 注意到要使得 D 等于 k, k 一定要同时整除 X 和 Y, 且 X/k 和 Y/k 要互素 (即两数的最大公约数为 1).

(b) 利用 (a), 证明

$$Q_1 = P\{X\text{和}Y\ \text{互素}\} = \frac{1}{\sum\limits_{k=1}^{\infty} 1/k^2}$$

根据著名的等式 $\sum_1^\infty 1/k^2 = \pi^2/6$, 因此 $Q_1 = 6/\pi^2$.(在数论里, 这就是著名的勒让德定理.)

(c) 证明

$$Q_1 = \prod_{i=1}^{\infty}\left(\frac{P_i^2-1}{P_i^2}\right)$$

其中 P_i 是大于 1 的第 i 个最小的素数.

提示: X 和 Y 互素的条件是它们没有公共的素因子, 因此, 利用 (b), 可以看出

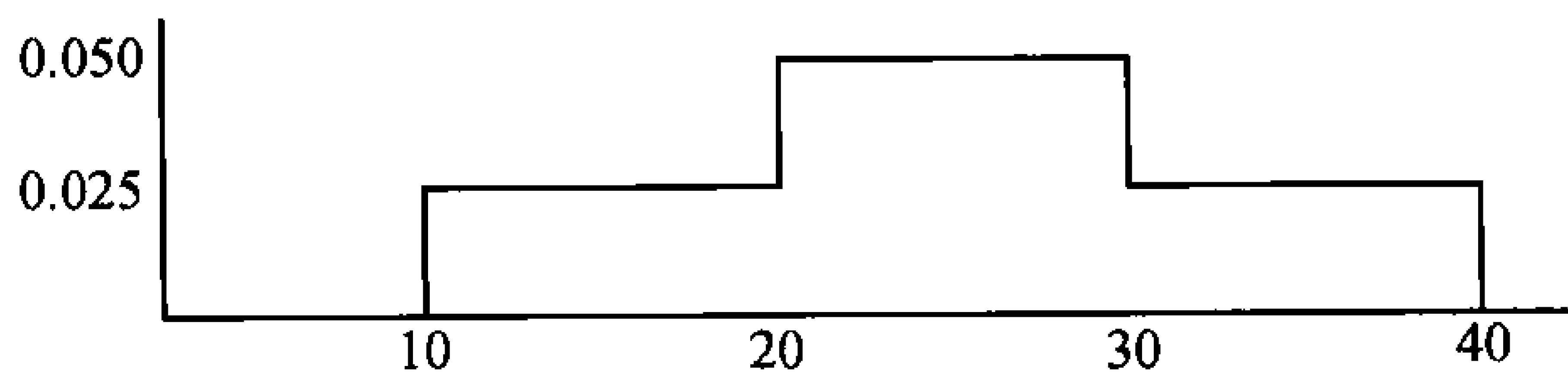
$$\prod_{i=1}^{\infty}\left(\frac{P_i^2-1}{P_i^2}\right) = \frac{6}{\pi^2}$$

这点在第 4 章的习题 11 中提及过. (本题与第 4 章习题 11 之间的关系就是, X 和 Y 互素的条件是 X 和 Y 没有公共的素因子.)

33. 当 $g(x)$ 是递减函数时, 证明定理 7.1.

自检习题

1. 某高中篮球队员在随机挑的一场篮球比赛中的上场时间是一个随机变量, 其概率密度函数如下图给出:



求以下事件概率:

(a) 上场时间超过 15 分钟;　(b) 上场时间在 20 到 35 分钟之间;

(c) 上场时间小于 30 分钟;　(d) 上场时间超过 36 分钟.

2. 随机变量 X 的密度函数为

$$f(x) = \begin{cases} cx^n & 0 < x < 1 \\ 0 & \text{其他} \end{cases}$$

其中 c 为常数, 计算 (a) c, (b) $P\{X > x\}, 0 < x < 1$.

3. 随机变量 X 的概率密度函数为

$$f(x) = \begin{cases} cx^4 & 0 < x < 2 \\ 0 & \text{其他} \end{cases}$$

其中 c 为常数, 计算 (a) $E[X]$, (b) $\text{Var}(X)$.

4. 随机变量 X 的概率密度函数为

$$f(x)=\begin{cases}ax+bx^2 & 0<x<1\\ 0 & \text{其他}\end{cases}$$

如果 $E[X]=0.6$, 计算 (a) $P\{X<1/2\}$, (b) $\mathrm{Var}(X)$.

5. 随机变量 X 称为取值为整数 $1,2,\cdots,n$ 的离散型均匀分布的随机变量, 如果

$$P\{X=i\}=1/n \qquad i=1,2,\cdots,n$$

对于任意非负实数 x, 记 $\mathrm{Int}(x)$(有时也记为 $[x]$) 为不超过 x 的最大整数. 如果 U 是一个服从 $(0,1)$ 上的均匀分布的随机变量, 证明 $X=\mathrm{Int}(nU)+1$ 为取值为 $1,2,\cdots,n$ 的离散型均匀分布的随机变量.

6. 在某个竞标工程里, 你们公司将要做出竞价. 如果你赢得了这个合约 (出价最低), 那么你计划还要付出 100(单位: 千美元) 完成这个工程. 如果你认为其他竞标者的最低竞价服从 $(70,140)$ 上的均匀分布, 那么你应该出价多少使得你的期望赢利最大?

7. 在如下游戏中, 你为了取胜, 需要连续胜 3 局. 该游戏决定于一个 $(0,1)$ 上均匀分布的值, 如果 $U>0.1$, 那么你第 1 局获胜; 如果 $U>0.2$, 那么你第二局获胜; 如果 $U>0.3$, 那么你第三局获胜.

   (a) 计算赢得第一局的概率;
   (b) 计算赢得第一局的条件下, 赢得第二局的条件概率;
   (c) 计算赢得第一局和第二局的条件下, 赢得第三局的条件概率;
   (d) 计算你最后获胜的概率.

8. 随机选中的一名参加 IQ 测试者得到的智商值近似服从均值为 100, 标准差为 15 的正态分布. 求以下事件概率:

   (a) 该测试者的智商值在 125 以上; (b) 该测试者的智商值在 90 到 110 之间.

9. 假设从您家到办公室所要花费时间服从均值为 40 分钟, 标准差为 7 分钟的正态分布. 如果你要参加一个下午 1 点的约会, 地点就在办公室, 那么要想做到 95% 的把握不迟到, 那么最晚何时出发?

10. 某汽车轮胎的寿命服从均值为 34 000 英里, 标准差为 4000 英里的正态分布.

    (a) 该轮胎寿命超过 40 000 英里的概率;
    (b) 该轮胎的寿命在 30 000 到 35 000 英里的概率;
    (c) 在该轮胎已经行驶了 30 000 英里的条件下, 那么还能行驶 10 000 英里的条件概率是多大?

11. 俄亥俄州克利夫兰市的年降雨量近似为均值为 40.2 英寸, 标准差为 8.4 英寸的正态随机变量. 求以下事件概率:

    (a) 下一年降雨量超过 44 英寸;
    (b) 接下来的 7 年中, 正好有 3 年的降雨量超过 44 英寸.

    假定 $A_i, i\geqslant 1$ 相互独立, 其中 A_i 表示“接下来的第 i 年的降雨量超过 44 英寸”.

12. 下表是 1992 年男女职工的收入表：

收入范围	女性百分比	男性百分比
⩽ 9999	8.6	4.4
10 000～19 999	38.0	21.1
20 000～24 999	19.4	15.8
25 000～49 999	29.2	41.5
⩾50 000	4.8	17.2

考虑随机抽取 200 个男职工和 200 个女职工，估算以下概率的近似值：

(a) 至少 70 个女职工的收入超过 25 000； (b) 至少 60% 的男职工收入超过 25 000；

(b) 至少 3/4 的男职工和至少一半的女职工收入超过 20 000.

13. 在某个银行，顾客办理业务的时间的长度服从均值为 5 分钟的指数分布，如果你走进银行时，正好有一个顾客在办理业务，问他或她还要再办理 4 分钟的概率是多大？

14. 假设随机变量的分布函数如下：

$$F(x) = 1 - \mathrm{e}^{-x^2} \qquad x > 0$$

求：

(a) $P\{X > 2\}$； (b) $P\{1 < X < 3\}$； (c) F 的危险率函数；

(d) $E[X]$； (e) $\mathrm{Var}(X)$.

提示：(d) 和 (e) 需要利用理论习题 5 的结论.

15. 洗衣机工作的时间是一个随机变量 (单位：年)，其失效率函数如下：

$$\lambda(t) = \begin{cases} 0.2 & 0 < t < 2 \\ 0.2 + 0.3(t-2) & 2 \leqslant t < 5 \\ 1.1 & t > 5 \end{cases}$$

(a) 购买 6 年后，洗衣机仍能正常工作的概率是多大；

(b) 在购买 6 年后，洗衣机接下来 2 年内会损坏的概率是多大？

16. 标准柯西分布随机变量的概率密度函数为

$$f(x) = \frac{1}{\pi(1+x^2)} \qquad -\infty < x < \infty$$

如果 X 为标准柯西随机变量，证明 $1/X$ 也是标准柯西随机变量.

17. 轮盘赌有 38 个下注区，分别标有 0, 00, 1 到 36. 如果你押 1 元在某个数字上，那么当轮盘停在该数字上时，你能赢得 35 元，否则就输掉这 1 元. 如果你连续进行这样的下注，求以下概率的近似值：

(a) 在 34 次下注后，您总的来说是获利的；

(b) 在 1000 次下注后，您总的来说是获利的；

(c) 在 100 000 次下注后，您总的来说是获利的.

假定每次轮盘转动以后, 将等概率地停留在这 38 个可能的注区上.

18. 箱子里有两种电池, 第 i 种使用的时间分别服从参数为 $\lambda_i, i = 1, 2$ 的指数分布. 随机地从中取一个电池, 取到第 i 种的概率为 p_i, $\sum_{i=1}^{2} p_i = 1$. 如果随机取出的一个电池使用 t 小时后仍正常使用, 那么它还能继续使用 s 小时的概率是多大?

19. 涉及犯罪嫌疑人的证据是一个随机变量 X. X 为一指数型随机变量, 其均值为 μ. 若该人无罪, 则 $\mu = 1$, 否则 $\mu = 2$. 法官按下列方式判罪: 当 $X > c$ 时判有罪, 否则判无罪.

(a) 法官希望以 95% 的把握不冤枉一个无罪的人, c 应该取多少?

(b) 利用 (a) 中得到的 c 值, 计算将一个确实有罪的被告判为有罪的概率.

20. 对任何实数 y, 定义

$$y^+ = \begin{cases} y, & y \geqslant 0 \\ 0, & y < 0 \end{cases}$$

令 c 为常数.

(a) 证明 $E[(Z-c)^+] = \dfrac{1}{\sqrt{2\pi}} e^{-c^2/2} - c(1 - \Phi(c))$, 其中 Z 为标准正态随机变量.

(b) 求 $E[(X-c)^+]$, 其中 X 为正态随机变量, 均值为 μ, 方差为 σ^2.

第 6 章　随机变量的联合分布

6.1　联合分布函数

到现在为止，我们仅仅探讨了单一随机变量的概率分布，然而，我们经常还对两个或两个以上的的随机变量的有关概率问题感兴趣. 为了处理类似问题，我们定义任意两个随机变量 X 和 Y 的**联合分布函数**(joint cummulative probability distribution function) 如下：

$$F(a,b)=P\{X\leqslant a,Y\leqslant b\}\qquad -\infty<a,b<\infty$$

X 的分布可以通过 X 和 Y 的联合分布得到：

$$\begin{aligned}F_X(a)&=P\{X\leqslant a\}=P\{X\leqslant a,Y<\infty\}\\&=P\big(\lim_{b\to\infty}\{X\leqslant a,Y\leqslant b\}\big)=\lim_{b\to\infty}P\{X\leqslant a,Y\leqslant b\}\\&=\lim_{b\to\infty}F(a,b)\equiv F(a,\infty)\end{aligned}$$

读者会注意到，在上述一系列等式里，我们又一次用到了概率是一个连续的集 (事件) 函数这一事实. 类似，Y 的分布函数如下：

$$F_Y(b)=P\{Y\leqslant b\}=\lim_{a\to\infty}F(a,b)\equiv F(\infty,b)$$

分布函数 F_X 和 F_Y 称为 X 和 Y 的**边缘分布**(marginal distribution).

理论上，所有有关 X 和 Y 的概率问题都可以通过其联合分布函数来解决. 比如，如果需要知道 $X>a$ 和 $Y>b$ 的联合概率，那么可以如下计算：

$$\begin{aligned}P\{X>a,Y>b\}&=1-P\big(\{X>a,Y>b\}^{c}\big)\\&=1-P\big(\{X>a\}^{c}\cup\{Y>b\}^{c}\big)=1-P\big(\{X\leqslant a\}\cup\{Y\leqslant b\}\big)\\&=1-\big[P\{X\leqslant a\}+P\{Y\leqslant b\}-P\{X\leqslant a,Y\leqslant b\}\big]\\&=1-F_X(a)-F_Y(b)+F(a,b)\end{aligned}\tag{1.1}$$

公式 (1.1) 是以下公式 (证明留作习题) 的特例：

$$\begin{aligned}&P\{a_1<X\leqslant a_2,b_1<Y\leqslant b_2\}\\=&F(a_2,b_2)+F(a_1,b_1)-F(a_1,b_2)-F(a_2,b_1)\end{aligned}\tag{1.2}$$

其中 $a_1<a_2,b_1<b_2$.

当 X 和 Y 都是离散型随机变量时，X 和 Y 的**联合分布列**(joint probability

mass function) 可以这样方便地定义:

$$p(x, y) = P\{X = x, Y = y\}$$

通过 $p(x, y)$ 可以得到 X 的分布列:

$$p_X(x) = P\{X = x\} = \sum_{y:p(x,y)>0} p(x, y)$$

类似地,

$$p_Y(y) = \sum_{x:p(x,y)>0} p(x, y)$$

例 1a 坛子里有 3 个红球, 4 个白球, 及 5 个蓝球, 从中随机取 3 个球. 令 X 和 Y 分别表示取出的红球数和白球数, 那么 X 和 Y 的联合分布列 $p(i, j) = P\{X = i, Y = j\}$ 计算如下:

$$p(0,0) = \binom{5}{3}\Big/\binom{12}{3} = \frac{10}{220} \qquad p(0,1) = \binom{4}{1}\binom{5}{2}\Big/\binom{12}{3} = \frac{40}{220}$$

$$p(0,2) = \binom{4}{2}\binom{5}{1}\Big/\binom{12}{3} = \frac{30}{220} \qquad p(0,3) = \binom{4}{3}\Big/\binom{12}{3} = \frac{4}{220}$$

$$p(1,0) = \binom{3}{1}\binom{5}{2}\Big/\binom{12}{3} = \frac{30}{220} \qquad p(1,1) = \binom{3}{1}\binom{4}{1}\binom{5}{1}\Big/\binom{12}{3} = \frac{60}{220}$$

$$p(1,2) = \binom{3}{1}\binom{4}{2}\Big/\binom{12}{3} = \frac{18}{220} \qquad p(2,0) = \binom{3}{2}\binom{5}{1}\Big/\binom{12}{3} = \frac{15}{220}$$

$$p(2,1) = \binom{3}{2}\binom{4}{1}\Big/\binom{12}{3} = \frac{12}{220} \qquad p(3,0) = \binom{3}{3}\Big/\binom{12}{3} = \frac{1}{220}$$

这些概率也可简单表示成表 6.1 的形式. 读者会注意到, X 的分布列可以对行求和得到, 而 Y 的分布列可以通过对列求和得到. 既然 X 和 Y 各自的分布列都出现在这样表格的边缘, 因此, 它们又常常分别称为 X 和 Y 的的边缘分布列. ■

表 6.1 $P\{X = i, Y = j\}$

i \ j	0	1	2	3	行和: $P\{X = i\}$
0	$\frac{10}{220}$	$\frac{40}{220}$	$\frac{30}{220}$	$\frac{4}{220}$	$\frac{84}{220}$
1	$\frac{30}{220}$	$\frac{60}{220}$	$\frac{18}{220}$	0	$\frac{108}{220}$
2	$\frac{15}{220}$	$\frac{12}{220}$	0	0	$\frac{27}{220}$
3	$\frac{1}{220}$	0	0	0	$\frac{1}{220}$
列和: $P\{Y = j\}$	$\frac{56}{220}$	$\frac{112}{220}$	$\frac{48}{220}$	$\frac{4}{220}$	

例 1b 假设某个社区内，15% 的家庭没有小孩，20% 的家庭有一个小孩，35% 的家庭有两个小孩，30% 的家庭有 3 个小孩．而且，进一步假定每个家庭里的每个孩子为男孩或女孩的可能性是一样的 (且独立)．如果从这个社区内随机抽取一个家庭，令 B 表示这个家庭的男孩数，令 G 表示该家庭里女孩数，那么它们的联合分布列如表 6.2 所示．

由这些概率得到：

$$P\{B=0,G=0\}=P\{\text{没有孩子}\}=0.15$$

$$\begin{aligned}P\{B=0,G=1\}&=P\{1\text{ 个孩子且为女孩}\}\\&=P\{1\text{个孩子}\}P\{1\text{个女孩}|1\text{个孩子}\}=0.20\times\frac{1}{2}\end{aligned}$$

$$\begin{aligned}P\{B=0,G=2\}&=P\{\text{ 2 个孩子且都为女孩}\}\\&=P\{2\text{个孩子}\}P\{2\text{个女孩}|2\text{个孩子}\}=0.35\times\left(\frac{1}{2}\right)^2\end{aligned}$$

表 6.2 中其余概率的验证留给读者． ■

表 6.2 $P\{B=i,G=j\}$

i \ j	0	1	2	3	行和：$P\{B=i\}$
0	0.15	0.10	0.0875	0.0375	0.3750
1	0.10	0.175	0.1125	0	0.3875
2	0.0875	0.1125	0	0	0.2000
3	0.0375	0	0	0	0.0375
列和：$P\{G=j\}$	0.3750	0.3875	0.2000	0.0375	

我们称 X 和 Y 是**联合连续**(jointly continuous) 的，如果存在一个对任意 x,y 定义的函数 $f(x,y)$，有以下性质：对任意实数对集合 C(也即 C 是两维空间里的集合)，有

$$P\{(X,Y)\in C\}=\iint\limits_{(x,y)\in C}f(x,y)\,\mathrm{d}x\,\mathrm{d}y \tag{1.3}$$

函数 $f(x,y)$ 称为 X 和 Y 的**联合密度函数**(joint probability density function)．如果 A 和 B 为任意实数集，定义 $C=\{(x,y):x\in A,y\in B\}$，通过公式 (1.3) 可以看出：

$$P\{X\in A,Y\in B\}=\int_B\int_A f(x,y)\,\mathrm{d}x\,\mathrm{d}y \tag{1.4}$$

因为

$$F(a,b)=P\{X\in(-\infty,a),Y\in(-\infty,b)\}=\int_{-\infty}^{b}\int_{-\infty}^{a}f(x,y)\,\mathrm{d}x\,\mathrm{d}y$$

只要偏导数有定义就可得：

$$f(a,b)=\frac{\partial^2}{\partial a\,\partial b}F(a,b)$$

可以从另一个角度来理解连续密度函数的定义, 利用 (1.4) 式, 可得到公式

$$P\{a<X<a+\mathrm{d}a, b<Y<b+\mathrm{d}b\}=\int_b^{b+\mathrm{d}b}\int_a^{a+\mathrm{d}a}f(x,y)\,\mathrm{d}x\,\mathrm{d}y\approx f(a,b)\,\mathrm{d}a\,\mathrm{d}b$$

其中 $\mathrm{d}a$ 和 $\mathrm{d}b$ 很小, 且 $f(x,y)$ 在 (a,b) 处连续. 因此, $f(a,b)$ 表示为随机向量 (X,Y) 取值于 (a,b) 附近的可能性.

如果 X 和 Y 为联合连续的, 则它们各自都连续. 它们各自的密度函数可以如下得到:

$$P\{X\in A\}=P\{X\in A, Y\in(-\infty,\infty)\}=\int_A\int_{-\infty}^{\infty}f(x,y)\,\mathrm{d}y\,\mathrm{d}x=\int_A f_X(x)\,\mathrm{d}x$$

其中

$$f_X(x)=\int_{-\infty}^{\infty}f(x,y)\,\mathrm{d}y$$

是 X 的密度函数. 类似地, Y 的密度函数如下:

$$f_Y(y)=\int_{-\infty}^{\infty}f(x,y)\,\mathrm{d}x$$

例 1c 设 X 和 Y 的联合密度函数为:

$$f(x,y)=\begin{cases}2\mathrm{e}^{-x}\mathrm{e}^{-2y} & 0<x<\infty, 0<y<\infty\\ 0 & \text{其他}\end{cases}$$

计算: (a) $P\{X>1, Y<1\}$; (b) $P\{X<Y\}$; (c) $P\{X<a\}$.

解: (a)

$$\begin{aligned}P\{X>1, Y<1\}&=\int_0^1\int_1^{\infty}2\mathrm{e}^{-x}\mathrm{e}^{-2y}\,\mathrm{d}x\,\mathrm{d}y\\&=\int_0^1 2\mathrm{e}^{-2y}(-\mathrm{e}^{-x}|_1^{\infty})\,\mathrm{d}y=\mathrm{e}^{-1}\int_0^1 2\mathrm{e}^{-2y}\,\mathrm{d}y=\mathrm{e}^{-1}(1-\mathrm{e}^{-2})\end{aligned}$$

(b)

$$\begin{aligned}P\{X<Y\}&=\iint\limits_{(x,y):x<y}2\mathrm{e}^{-x}\mathrm{e}^{-2y}\,\mathrm{d}x\,\mathrm{d}y=\int_0^{\infty}\int_0^{y}2\mathrm{e}^{-x}\mathrm{e}^{-2y}\,\mathrm{d}x\,\mathrm{d}y\\&=\int_0^{\infty}2\mathrm{e}^{-2y}(1-\mathrm{e}^{-y})\mathrm{d}y=\int_0^{\infty}2\mathrm{e}^{-2y}\,\mathrm{d}y-\int_0^{\infty}2\mathrm{e}^{-3y}\,\mathrm{d}y=1-\frac{2}{3}=\frac{1}{3}\end{aligned}$$

(c)

$$P\{X<a\}=\int_0^a\int_0^{\infty}2\mathrm{e}^{-2y}\mathrm{e}^{-x}\,\mathrm{d}y\,\mathrm{d}x=\int_0^a\mathrm{e}^{-x}\,\mathrm{d}x=1-\mathrm{e}^{-a}$$

■

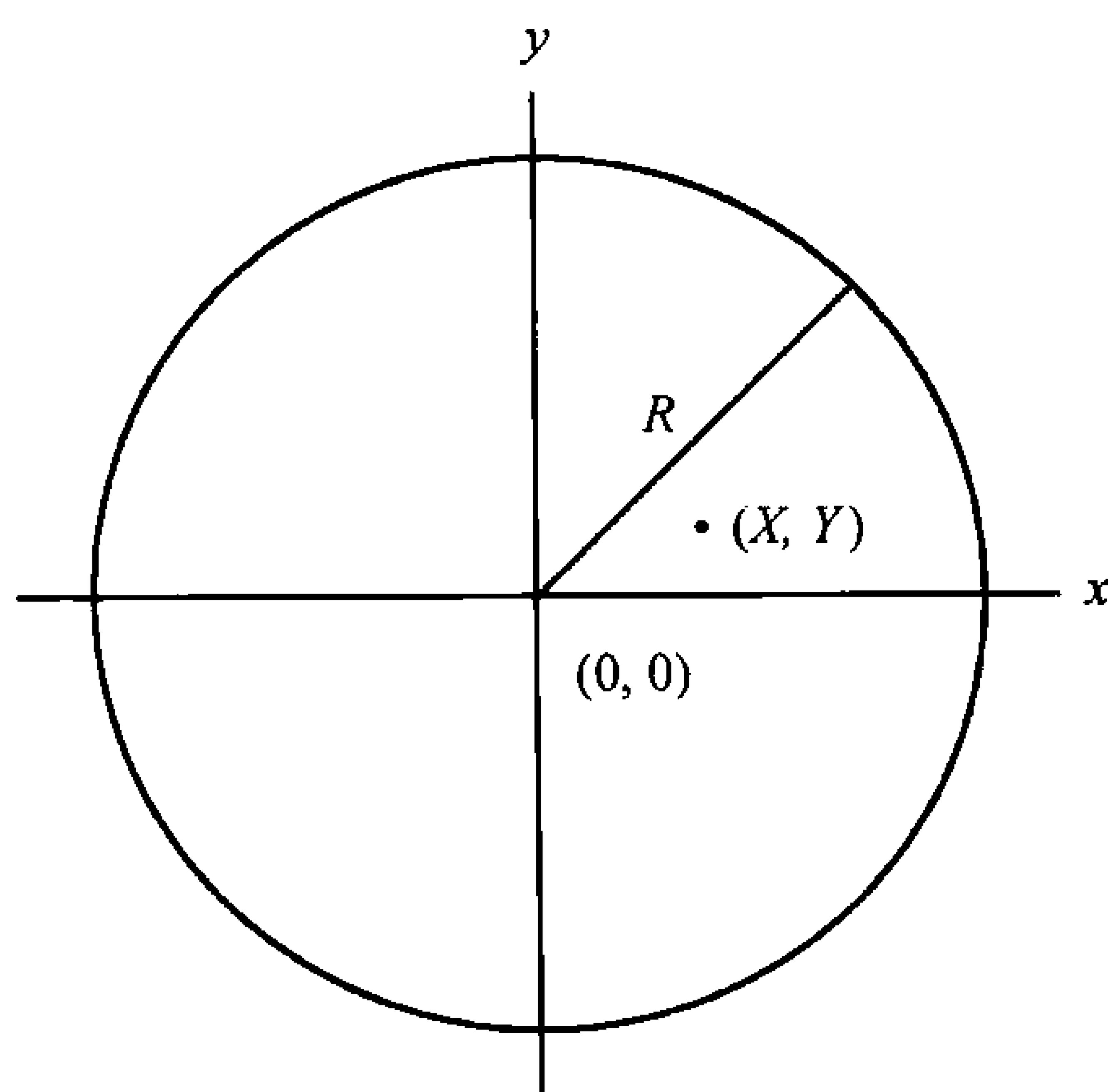


图 6.1 联合概率分布

例 1d 考虑一个半径为 R 的圆, 按如下方式随机从圆内选一点: 落在圆内任一区域内的概率只与这个区域的面积有关, 与该区域在圆内位置无关 (也即该点在圆内服从均匀分布). 如果令圆心表示原点, 且令 X 和 Y 表示该点的坐标 (如图 6.1). 既然 (X,Y) 落在圆内任一点附近的概率都是一样的, X 和 Y 的联合密度函数为

$$f(x,y)=\begin{cases}c & \text{如果}x^2+y^2\leqslant R^2\\ 0 & \text{如果}x^2+y^2>R^2\end{cases}$$

(a) 求常数 c; (b) 计算 X 和 Y 的边缘密度函数;

(c) 求原点到该点的距离 D 小于等于 a 的概率; (d) 计算 $E[D]$.

解: (a) 利用密度的性质

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)\,\mathrm{d}y\,\mathrm{d}x=1$$

可得

$$c\iint\limits_{x^2+y^2\leqslant R^2}\mathrm{d}y\,\mathrm{d}x=1$$

我们可以利用极坐标计算 $\iint_{x^2+y^2\leqslant R^2}\mathrm{d}y\,\mathrm{d}x$ 的值. 也可以用更简单的办法, 注意到它表示的就是半径为 R 的圆的面积, 因此等于 πR^2, 从而

$$c=\frac{1}{\pi R^2}$$

(b)

$$\begin{aligned}f_X(x)&=\int_{-\infty}^{\infty}f(x,y)\,\mathrm{d}y=\frac{1}{\pi R^2}\int_{x^2+y^2\leqslant R^2}\mathrm{d}y\\&=\frac{1}{\pi R^2}\int_{-c}^{c}\mathrm{d}y\qquad\text{其中}c=\sqrt{R^2-x^2}\\&=\frac{2}{\pi R^2}\sqrt{R^2-x^2}\qquad x^2\leqslant R^2\end{aligned}$$

当 $x^2>R^2$ 时, 它等于 0. 利用对称性可知, Y 的边缘密度为

$$f_Y(y)=\begin{cases}\dfrac{2}{\pi R^2}\sqrt{R^2-y^2} & y^2\leqslant R^2\\ 0 & y^2>R^2\end{cases}$$

(c) 原点到该点的距离 $D=\sqrt{X^2+Y^2}$ 的分布函数可以如下得到, 对于 $0\leqslant$

$a \leqslant R$, 有

$$F_D(a) = P\{\sqrt{X^2+Y^2} \leqslant a\} = P\{X^2+Y^2 \leqslant a^2\}$$
$$= \iint_{x^2+y^2 \leqslant a^2} f(x,y)\,\mathrm{d}y\,\mathrm{d}x = \frac{1}{\pi R^2}\iint_{x^2+y^2 \leqslant a^2} \mathrm{d}y\,\mathrm{d}x = \frac{\pi a^2}{\pi R^2} = \frac{a^2}{R^2}$$

其中用到了事实 $\iint_{x^2+y^2\leqslant a^2} \mathrm{d}y\,\mathrm{d}x$ 为半径为 a 的圆的面积, 其值为 πa^2.

(d) 从以上 (c) 可以得到 D 的密度函数为

$$f_D(a) = \frac{2a}{R^2} \qquad 0 \leqslant a \leqslant R$$

因此,

$$E[D] = \frac{2}{R^2}\int_0^R a^2\,\mathrm{d}a = \frac{2R}{3}$$

■

例 1e X 和 Y 的联合密度为

$$f(x,y) = \begin{cases} \mathrm{e}^{-(x+y)} & 0 < x < \infty, \quad 0 < y < \infty \\ 0 & \text{其他} \end{cases}$$

求随机变量 X/Y 的密度函数.

解: 先来计算 X/Y 的分布函数. 对于 $a > 0$, 有

$$F_{X/Y}(a) = P\left\{\frac{X}{Y} \leqslant a\right\} = \iint_{x/y \leqslant a} \mathrm{e}^{-(x+y)}\,\mathrm{d}x\,\mathrm{d}y = \int_0^\infty \int_0^{ay} \mathrm{e}^{-(x+y)}\,\mathrm{d}x\,\mathrm{d}y$$
$$= \int_0^\infty (1-\mathrm{e}^{-ay})\mathrm{e}^{-y}\,\mathrm{d}y = \left\{-\mathrm{e}^{-y} + \frac{\mathrm{e}^{-(a+1)y}}{a+1}\right\}\Big|_0^\infty = 1 - \frac{1}{a+1}$$

对 $F_{X/Y}(a)$ 求导可得到 X/Y 的密度函数 $f_{X/Y}(a) = 1/(a+1)^2, 0 < a < \infty$. ■

我们也可以用和 $n=2$ 时同样的方法定义 n 个随机变量的联合分布. 比如, n 个随机变量 $X_1, X_2, \cdots, X_n$ 的联合分布函数定义为

$$F(a_1, a_2, \cdots, a_n) = P\{X_1 \leqslant a_1, X_2 \leqslant a_2, \cdots, X_n \leqslant a_n\}$$

而且, $X_1, \cdots, X_n$ 称为联合连续的, 如果存在一个函数 $f(x_1, \cdots, x_n)$, 对于 n 维空间里的任意集合 C, 满足下列条件:

$$P\{(X_1, X_2, \cdots, X_n) \in C\} = \int\int_{(x_1,\cdots,x_n)\in C}\cdots\int f(x_1, \cdots, x_n)\,\mathrm{d}x_1\,\mathrm{d}x_2 \cdots \mathrm{d}x_n$$

$f(x_1, \cdots, x_n)$称为$X_1, \cdots, X_n$的联合密度函数. 特别, 对于 n 个实数集$A_1, A_2, \cdots, A_n$,

有

$$P\{X_1 \in A_1, X_2 \in A_2, \cdots, X_n \in A_n\}$$
$$= \int_{A_n}\int_{A_{n-1}}\cdots\int_{A_1} f(x_1, x_2, \cdots, x_n)\,\mathrm{d}x_1\,\mathrm{d}x_2\cdots\mathrm{d}x_n$$

例 1f (多项分布) 多项分布是很重要的联合分布之一, 它经常出现在进行 n 次独立重复试验当中. 假设每次试验都有 r 种可能结果, 各自概率分别为 $p_1, p_2, \cdots, p_r, \sum_{i=1}^r p_i = 1$, 令 X_i 表示 n 次试验中第 i 个结果出现的次数, 那么

$$P\{X_1 = n_1, X_2 = n_2, \cdots, X_r = n_r\} = \frac{n!}{n_1!n_2!\cdots n_r!}p_1^{n_1}p_2^{n_2}\cdots p_r^{n_r} \tag{1.5}$$

其中 $\sum_{i=1}^r n_i = n, n_i \geqslant 0, i = 1, \cdots, r$.

现在来解释 (1.5) 式. 记 n 次试验结果为 $(i_1, i_2, \cdots, i_n)$, 其中 i_j 取值于 $\{1, 2, \cdots, r\}$ (这里我们记数字 i 表示第 i 种结果 $i = 1, \cdots, r$), 表示第 j 次试验的试验结果. 若 $i_1, \cdots, i_n$ 中有 n_1 个 $1, \cdots, n_r$ 个 r, 则由试验的独立性可知

$$P\{(i_1, \cdots, i_n)\} = p_1^{n_1}\cdots p_r^{n_r}$$

事件 $\{X_1 = n_1, \cdots, X_r = n_r\}$ 可作如下分解:

$$\{X_1 = n_1, \cdots, X_r = n_r\} = \bigcup\{(i_1, \cdots, i_n)\}$$

其中事件求和号是表示对一切 $(i_1, \cdots, i_n)$ 求和, 其中 $(i_1, \cdots, i_n)$ 中有 n_1 个 $1, \cdots, n_r$ 个 r. 利用概率的性质可知

$$P\{X_1 = n_1, \cdots, X_r = n_r\} = \sum P\{(i_1, \cdots, i_n)\} = \sum p_1^{n_1}\cdots p_r^{n_r}$$

显然求和的项数为 $n!/(n_1!\cdots n_r!)$. 这样, 我们得到 (1.5) 式. 以 (1.5) 式为联合分布列的联合分布称为*多项分布*. 注意, 当 $r=2$ 时, 多项分布就退化为二项分布.

再次说明, 对 $\{X_1, X_2, \cdots, X_r\}$ 的任何子集求和, 这个和都具有二项分布. 也就是说, 若 $N \subset \{1, 2, \cdots, r\}$, 则 $\sum_{i\in N} X_i$ 将是二项随机变量, 参数为 n 和 $P = \sum_{i\in N} P_i$. 这是因为 $\sum_{i\in N} X_i$ 表示 n 次试验中, 结果属于 N 的试验次数, 每个试验独立且出现这个结果的概率是 $\sum_{i\in N} P_i$.

作为多项分布的应用, 考虑掷一枚均匀的骰子 9 次, 那么 1 出现 3 次, 2 和 3 各出现 2 次, 4 和 5 各出现 1 次, 6 不出现的概率为

$$\frac{9!}{3!2!2!1!1!0!}\left(\frac{1}{6}\right)^3\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^1\left(\frac{1}{6}\right)^1\left(\frac{1}{6}\right)^0 = \frac{9!}{3!2!2!}\left(\frac{1}{6}\right)^9$$ ■

6.2 独立随机变量

随机变量 X 和 Y 称为*独立的* (independent), 如果对任意两个实数集 A 和 B,

有

$$P\{X \in A, Y \in B\} = P\{X \in A\}P\{Y \in B\} \tag{2.1}$$

也就是说, 如果对所有的 A 和 B, 事件 $E_A = \{X \in A\}$ 和 $F_B = \{Y \in B\}$ 是独立的, 那么随机变量 X 和 Y 独立.

利用概率的三条公理可知, 公式 (2.1) 成立当且仅当对所有 a, b, 有

$$P\{X \leqslant a, Y \leqslant b\} = P\{X \leqslant a\}P\{Y \leqslant b\}$$

因此, 利用 X 和 Y 的联合分布函数 F, 我们有, 如果

$$F(a,b) = F_X(a)F_Y(b) \qquad \text{对所有的 } a,b \text{ 成立}$$

则 X 和 Y 独立. 当 X 和 Y 为离散型随机变量时, 独立的条件 (2.1) 等价于

$$p(x,y) = p_X(x)p_Y(y) \qquad \text{对所有的 } x,y \tag{2.2}$$

上述结论成立是因为如果 (2.1) 成立, 那么令 A 和 B 分别表示单点集 $A=\{x\}$ 和 $B=\{y\}$, (2.1) 式就变成了 (2.2) 式反之, 如果 (2.2) 式成立, 那么对任意集合 A, B, 有

$$\begin{aligned} P\{X \in A, Y \in B\} &= \sum_{y\in B}\sum_{x\in A} p(x,y) = \sum_{y\in B}\sum_{x\in A} p_X(x)p_Y(y) \\ &= \sum_{y\in B} p_Y(y) \sum_{x\in A} p_X(x) = P\{Y \in B\}P\{X \in A\} \end{aligned}$$

这样就证明了公式 (2.1) 式的成立.

在 X, Y 联合连续的情形下, 独立的条件等价于

$$f(x,y) = f_X(x)f_Y(y) \quad \text{对所有的 } x,y$$

因此, 粗略地说, 如果知道其中一个变量的取值并不影响另一个变量的分布, 则两个变量就相互独立. 不独立的随机变量称为*相依的*(dependent).

例 2a 考虑进行 $n+m$ 次独立重复试验, 每次成功的概率为 p, 如果 X 表示前 n 次试验成功的次数, Y 表示后 m 次试验成功的次数, 那么 X 和 Y 是独立的, 因为知道了前 n 次试验中成功的次数并不影响后 m 次试验中成功次数的分布 (因为假设试验是独立的). 事实上, 对于整数 x 和 y, 有

$$\begin{aligned} &P\{X = x, Y = y\} \\ &\quad = \binom{n}{x}p^x(1-p)^{n-x}\binom{m}{y}p^y(1-p)^{m-y} \qquad 0 \leqslant x \leqslant n, 0 \leqslant y \leqslant m \\ &\quad = P\{X = x\}P\{Y = y\} \end{aligned}$$

但是, X 和 Z 是相依的, 其中 Z 表示在 $n+m$ 次试验中总的成功次数 (为什么?). ■

例 2b 假设某一天内进入邮局的人数为服从参数为 λ 的泊松分布，证明：如果每个进入邮局的人为男性的概率为 p，为女性的概率为 $1-p$，则进入邮局的男人数和女人数是相互独立的泊松随机变量，它们的参数分别为 λp 和 $\lambda(1-p)$.

解：令 X 和 Y 分别表示进入邮局的男人数和女人数．为证明 X 和 Y 独立，只需证明 (2.2) 式成立．利用全概公式，

$$\begin{aligned}P\{X=i,Y=j\}=&P\{X=i,Y=j|X+Y=i+j\}P\{X+Y=i+j\}\\&+P\{X=i,Y=j|X+Y\neq i+j\}P\{X+Y\neq i+j\}\end{aligned}$$

(读者应该注意到该公式仅仅是公式 $P(E)=P(E|F)P(F)+P(E|F^c)P(F^c)$ 的特例.) 由于 $P\{X=i,Y=j|X+Y\neq i+j\}$ 显然为 0，我们可以得到

$$P\{X=i,Y=j\}=P\{X=i,Y=j|X+Y=i+j\}P\{X+Y=i+j\} \tag{2.3}$$

现在，因为 $X+Y$ 是进入邮局的总人数，根据假设

$$P\{X+Y=i+j\}=\mathrm{e}^{-\lambda}\frac{\lambda^{i+j}}{(i+j)!} \tag{2.4}$$

而且，在给定 $i+j$ 人进入邮局的情况下，既然每个进入邮局的人为男性的概率为 p，因此，正好有 i 个是男性 (且正好有 j 个女性) 的概率 $\binom{i+j}{i}p^i(1-p)^j$，即

$$P\{X=i,Y=j|X+Y=i+j\}=\binom{i+j}{i}p^i(1-p)^j \tag{2.5}$$

将公式 (2.4) 和 (2.5) 代入式子 (2.3)，可得

$$\begin{aligned}P\{X=i,Y=j\}&=\binom{i+j}{i}p^i(1-p)^j\mathrm{e}^{-\lambda}\frac{\lambda^{i+j}}{(i+j)!}\\&=\mathrm{e}^{-\lambda}\frac{(\lambda p)^i}{i!j!}[\lambda(1-p)]^j=\frac{\mathrm{e}^{-\lambda p}(\lambda p)^i}{i!}\mathrm{e}^{-\lambda(1-p)}\frac{[\lambda(1-p)]^j}{j!}\end{aligned} \tag{2.6}$$

因此

$$P\{X=i\}=\mathrm{e}^{-\lambda p}\frac{(\lambda p)^i}{i\,!}\sum_j \mathrm{e}^{-\lambda(1-p)}\frac{[\lambda(1-p)]^j}{j\,!}=\mathrm{e}^{-\lambda p}\frac{(\lambda p)^i}{i\,!} \tag{2.7}$$

类似地有

$$P\{Y=j\}=\mathrm{e}^{-\lambda(1-p)}\frac{[\lambda(1-p)]^j}{j!} \tag{2.8}$$

(2.6) 式、(2.7) 式和 (2.8) 式说明了所求结果. ■

例 2c 某男和某女决定在某个地点见面．如果每个人到达的时间是独立的，且在中午 12 点到下午 1 点之间均匀分布，求先到的人需要等待 10 分钟以上的概率．

解：令 X 和 Y 分别表示该男和该女到达的时间，以分钟为单位，以中午 12 点为起点．那么 X 和 Y 是相互独立的随机变量，且均服从 $(0,60)$ 上的均匀分布．所求概率为 $P\{X+10<Y\}+P\{Y+10<X\}$．根据对称性，它等于 $2P\{X+10<Y\}$.

如下得到:

$$2P\{X+10<Y\}=2\iint\limits_{x+10<y} f(x,y)\mathrm{d}x\,\mathrm{d}y=2\iint\limits_{x+10<y} f_X(x)f_Y(y)\,\mathrm{d}x\,\mathrm{d}y$$

$$=2\int_{10}^{60}\int_{0}^{y-10}\left(\frac{1}{60}\right)^2\mathrm{d}x\,\mathrm{d}y=\frac{2}{(60)^2}\int_{10}^{60}(y-10)\,\mathrm{d}y=\frac{25}{36}\quad ■$$

下面的例子是一个有关几何概率的古老问题. 它首先由 18 世纪的博物学家蒲丰提出并解决, 通常称为蒲丰投针问题.

例 2d (蒲丰投针问题) 桌面上画着一些平行线, 它们之间的距离都是 D, 向此桌面上随意投掷一长度为 L 的针, 其中 $L\leqslant D$, 问此针与桌面上的某一根平行线相交的概率是多大 (另一种可能是此针正好在某两条平行线之间)?

解: 从针的中点向距离该点最近的一条平行线引一条垂直线, 设这条垂线的长度为 X. 又设针与这条垂直线的夹角为 θ. 这样, 垂直线、平行线以及针所在直线会形成一个直角三角形 (见图 6.2). 如果直角三角形的斜边长小于 $L/2$ 时, 针会与这一条直线相交. 也即, 若

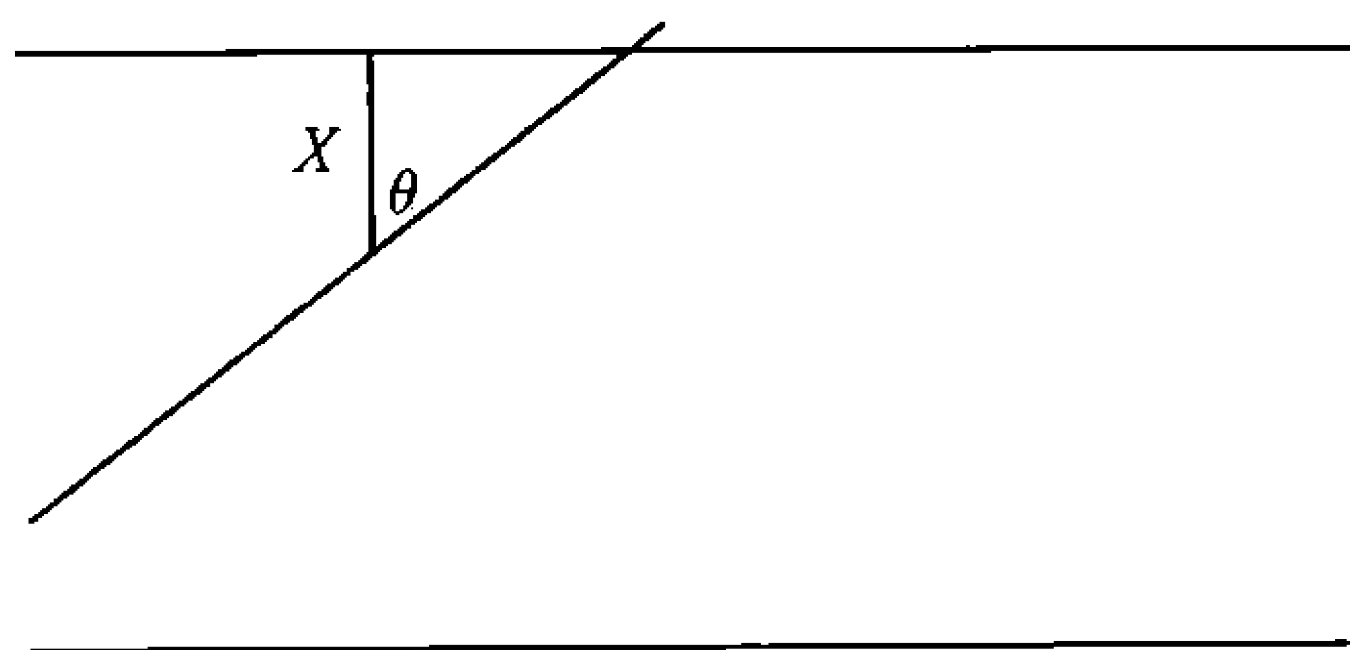


图 6.2 蒲丰投针

$$\frac{X}{\cos\theta}<\frac{L}{2}\quad 或\quad X<\frac{L}{2}\cos\theta$$

则针与这一条平行线相交. 注意 X 是一个取值于 0 到 $D/2$ 之间的随机变量, θ 是一个取值在 0 到 $\pi/2$ 之间的随机变量. 关于 X 和 θ, 很自然地假定它们是相互独立的, 并且在各自取值范围内均匀分布. 因此

$$P\left\{X<\frac{L}{2}\cos\theta\right\}=\iint\limits_{X<L/2\cos y} f_X(x)f_\theta(y)\,\mathrm{d}x\,\mathrm{d}y$$

$$=\frac{4}{\pi D}\int_0^{\pi/2}\int_0^{L/2\cos y}\mathrm{d}x\,\mathrm{d}y=\frac{4}{\pi D}\int_0^{\pi/2}\frac{L}{2}\cos y\,\mathrm{d}y=\frac{2L}{\pi D}\quad ■$$

例 2e (正态分布的特征) 令 X 和 Y 分别表示子弹的弹着点与靶心目标的水平和竖直偏差, 且假设

1. X 和 Y 为具有可微密度函数的独立的连续随机变量;

2. X 和 Y 的联合密度 $f(x,y)=f_X(x)f_Y(y)$ 作为 (x,y) 的函数只依赖 x^2+y^2 的值.

更直观地说来, 第 2 个假设说明了子弹落在 $x-y$ 平面的概率取决于弹着点与目标点的距离, 而与弹着点相对于目标的方位无关. 第 2 个假设的另一个等价的说法是联合密度函数相对于旋转是不变的.

下面的说法是令人感兴趣的: 由上述两个假设可推得 X 和 Y 为正态随机变

量. 为了证明这点, 首先注意到由假设可得

$$f(x,y)=f_X(x)f_Y(y)=g(x^2+y^2) \tag{2.9}$$

(2.9) 式两边对 x 求导, 可得

$$f'_X(x)f_Y(y)=2xg'(x^2+y^2) \tag{2.10}$$

用式 (2.9) 去除 (2.10), 得到

$$\frac{f'_X(x)}{f_X(x)}=\frac{2xg'(x^2+y^2)}{g(x^2+y^2)}$$

或

$$\frac{f'_X(x)}{2xf_X(x)}=\frac{g'(x^2+y^2)}{g(x^2+y^2)} \tag{2.11}$$

(2.11) 式的左边仅与 x 有关, 而 (2.11) 式的右边仅与 x^2+y^2 有关. 由此可以推出左边对任意 x 来说都是等值的. 为了证明这点, 考虑任意 x_1,x_2, 取 y_1,y_2 使其满足条件 $x_1^2+y_1^2=x_2^2+y_2^2$, 那么, 从 (2.11) 式可得

$$\frac{f'_X(x_1)}{2x_1f_X(x_1)}=\frac{g'(x_1^2+y_1^2)}{g(x_1^2+y_1^2)}=\frac{g'(x_2^2+y_2^2)}{g(x_2^2+y_2^2)}=\frac{f'_X(x_2)}{2x_2f_X(x_2)}$$

因此,

$$\frac{f'_X(x)}{xf_X(x)}=c \quad 或 \quad \frac{\mathrm{d}}{\mathrm{d}x}(\ln f_X(x))=cx$$

对两边积分可得

$$\ln f_X(x)=a+\frac{cx^2}{2} \quad 或 \quad f_X(x)=k\mathrm{e}^{cx^2/2}$$

由于 $\int_{-\infty}^{\infty}f_X(x)\mathrm{d}x=1$, c 必然为负数, 可将 c 写成 $c=-1/\sigma^2$, 即

$$f_X(x)=k\mathrm{e}^{-x^2/2\sigma^2}$$

也即, X 为一正态随机变量, 参数为 $\mu=0$ 和 σ^2. 类似地, 对于 $f_Y(y)$ 可证明

$$f_Y(y)=\frac{1}{\sqrt{2\pi}\bar{\sigma}}\mathrm{e}^{-y^2/2\bar{\sigma}^2}$$

再利用第二个假设可知 $\sigma^2=\bar{\sigma}^2$. 因此 X 和 Y 为独立同分布的正态随机变量, 其参数为 $\mu=0$ 和 σ^2. ■

X 和 Y 相互独立的一个充分且必要的条件是: 联合密度函数 (离散情形下为联合分布列) $f(x,y)$ 可以分解成两部分, 其中一部分仅与 x 有关, 另一部分仅与 y 有关.

命题 2.1 连续型 (离散型) 随机变量 X 和 Y 相互独立, 当且仅当其联合密度函数 (联合分布列) 可以写成

$$f_{X,Y}(x,y)=h(x)g(y) \qquad -\infty<x<\infty,-\infty<y<\infty$$

证明： 我们给出连续情形下的证明. 首先注意到独立意味着 X 和 Y 的联合密度函数等于其各自边缘密度函数的乘积, 因此, 当 X 和 Y 独立时, 上述因式分解是成立的. 现在, 假定 $f_{X,Y}(x,y)$ 具有下列分解式

$$f_{X,Y}(x,y)=h(x)g(y)$$

利用联合密度函数的性质可知

$$1=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y=\int_{-\infty}^{\infty}h(x)\,\mathrm{d}x\int_{-\infty}^{\infty}g(y)\,\mathrm{d}y=C_1C_2$$

其中 $C_1=\int_{-\infty}^{\infty}h(x)\,\mathrm{d}x$, $C_2=\int_{-\infty}^{\infty}g(y)\,\mathrm{d}y$. 这样

$$f_X(x)=\int_{-\infty}^{\infty}f_{X,Y}(x,y)\,\mathrm{d}y=C_2h(x)\qquad f_Y(y)=\int_{-\infty}^{\infty}f_{X,Y}(x,y)\,\mathrm{d}x=C_1g(y)$$

又由 $C_1C_2=1$, 可以得到

$$f_{X,Y}(x,y)=f_X(x)f_Y(y)$$

这样就完成了充分性的证明, 命题得证. □

例 2f 设 X 和 Y 的联合密度函数为

$$f(x,y)=\begin{cases}6\mathrm{e}^{-2x}\mathrm{e}^{-3y} & 0<x<\infty,0<y<\infty\\ 0 & \text{其他}\end{cases}$$

这两个随机变量是否独立? 如果联合密度函数如下呢?

$$f(x,y)=\begin{cases}24xy & 0<x<1,0<y<1,0<x+y<1\\ 0 & \text{其他}\end{cases}$$

解： 第一种情况下, 联合密度函数可分解因式, 因此随机变量 X 和 Y 相互独立 (其中一个为速率 $\lambda=2$, 另一个为速率 $\lambda=3$ 的指数分布). 第二种情况下, 因为联合密度函数非零的区域不能写成 $x\in A,y\in B$ 的形式, 联合密度函数不能进行因式分解, 因此, 随机变量 X 和 Y 并不独立. 通过如下方式更容易看出这一点, 令

$$I(x,y)=\begin{cases}1 & 0<x<1,0<y<1,0<x+y<1\\ 0 & \text{其他}\end{cases}$$

那么联合密度函数可写为

$$f(x,y)=24xyI(x,y)$$

很显然, 以上不能分解成分别仅与 x 和仅与 y 有关的两个因子. ■

当然, 对于两个以上的随机变量, 也可以给出独立性的定义. 一般来说, n 个随机变量 $X_1,X_2,\cdots,X_n$ 称为独立的, 如果以下条件满足: 对于任何实数集合 A_1, $A_2,\cdots,A_n$, 有

$$P\{X_1 \in A_1, X_2 \in A_2, \cdots, X_n \in A_n\} = \prod_{i=1}^{n} P\{X_i \in A_i\}$$

首先, 可以证明上述条件等价于

$$P\{X_1 \leqslant a_1, X_2 \leqslant a_2, \cdots, X_n \leqslant a_n\}$$
$$= \prod_{i=1}^{n} P\{X_i \leqslant a_i\} \qquad \text{对所有的 } a_1, a_2, \cdots, a_n$$

最后, 对于无限个随机变量的独立性可以这样定义的: 如果其中任意有限个随机变量都相互独立, 那么称这无限个随机变量是相互独立的.

例 2g (计算机怎样选择随机子集)[①] 大部分计算机都有内置程序, 用以产生或模拟 (0, 1) 上的均匀随机变量的值 (或称随机数). 作为应用, 计算机很容易模拟示性随机变量 (伯努利随机变量). 设 I 是一个示性随机变量, 即

$$P\{I = 1\} = p = 1 - P\{I = 0\}$$

计算机产生一随机数 U((0, 1) 上的均匀随机数), 令

$$I = \begin{cases} 1 & U < p \\ 0 & U \geqslant p \end{cases}$$

就模拟得到示性随机变量 I. 现在我们希望从 $\{1, 2, \cdots, n\}$ 中选择 k 个对象, 使得 $\binom{n}{k}$ 种不同的组合都有相同的机会被取上. 我们用 $(i_1, i_2, \cdots, i_n)$, 其中 $i_j = 0$或1(0 表示数字 j 没被选上, 1 则表示选上), $j = 1, 2, \cdots, n$, $\sum_{j=1}^{n} i_j = k$ 表示 k 个对象的一种选择结果. 令 $(I_1, \cdots, I_n)$ 表示 k 个对象的一个随机选择, 且使得 $\{1, 2, \cdots, n\}$ 的 $\binom{n}{k}$ 种不同的 k 个元素的组合都具有相同的机会被选中, 即 $(I_1, \cdots, I_n)$ 的分布列满足

$$P\{(I_1, \cdots, I_n) = (i_1, \cdots, i_n)\} = 1 \Big/ \binom{n}{k}$$

其中 $i_1, \cdots, i_n$ 取值于 $\{0, 1\}$, $\sum_{l=1}^{n} i_l = k$. 利用条件概率公式

$$\begin{aligned}
&P\{(I_1, \cdots, I_n) = (i_1, \cdots, i_n)\} \\
=&P\{I_1 = i_1\}P\{(I_2, \cdots, I_n) = (i_2, \cdots, i_n)|I_1 = i_1\} \\
=&P\{I_1 = i_1\}P\{I_2 = i_2|I_1 = i_1\}P\{(I_3, \cdots, I_n) = (i_3, \cdots, i_n)|I_1 = i_1, I_2 = i_2\} \\
&\cdots \\
=&P\{I_1 = i_1\}P\{I_2 = i_2|I_1 = i_1\}P\{I_3 = i_3|I_1 = i_1, I_2 = i_2\}\cdots \\
&P\{I_n = i_n|I_1 = i_1, \cdots, I_{n-1} = i_{n-1}\}
\end{aligned}$$

---

① 此例在讲解顺序上做了调整, 作者先讲述事件模拟的方法, 再在后面补充事件的概率的求法. 译者先说明事件的概率, 然后讲述模拟方法, 此法较符合大家的思维习惯. —— 译者注

其中, $I_1 = i_1$ 表示, 比如, 当 $i_1 = 1$ 时, 从集合 $\{1, 2\cdots, n\}$ 中随机地抽 k 个对象, 1 被抽中这一事件. 因此, $P\{I_1 = 1\} = k/n, P\{I_1 = 0\} = (n-k)/n$. 现在求条件分布 $P\{I_2 = i_2 | I_1 = i_1\}$, 有

$$P\{I_2 = 1 | I_1 = 1\} = \frac{k-1}{n-1} \qquad P\{I_2 = 0 | I_1 = 1\} = \frac{n-k}{n-1}$$

上式中第一个概率是 1 已经选中的情况下, 2 被选中的概率. 由于 1 已经选中, 剩下还需选 $k-1$ 个对象, 在 $\{2, \cdots, n\}$ 中 2 被选中的概率显然是 $(k-1)/(n-1)$. 综合以上分析, 对于 I_1, I_2 的分布, 我们有如下结论:

$$P\{I_1 = i_1\} = \begin{cases} \dfrac{k}{n} & i_1 = 1 \\ \dfrac{n-k}{n} & i_1 = 0 \end{cases} \qquad P\{I_2 = i_2 | I_1 = i_1\} = \begin{cases} \dfrac{k-i_1}{n-1} & i_2 = 1 \\ 1 - \dfrac{k-i_1}{n-1} & i_2 = 0 \end{cases}$$

对于一般的 l, 我们有

$$P\{I_{l+1} = i_{l+1} | I_1 = i_1, \cdots, I_l = i_l\} = \begin{cases} \dfrac{k - (i_1 + \cdots + i_l)}{n-l} & i_{l+1} = 1 \\ 1 - \dfrac{k - (i_1 + \cdots + i_l)}{n-l} & i_{l+1} = 0 \end{cases}$$

有了上述的条件概率公式, 我们就可以利用模拟的方法模拟上述 $I_1, \cdots, I_n$. 首先产生 n 个相互独立的 $(0,1)$ 上随机数 $U_1, \cdots, U_n$. 定义

$$I_1 = \begin{cases} 1 & U_1 < \dfrac{k}{n} \\ 0 & \text{其他} \end{cases}$$

$$I_2 = \begin{cases} 1 & U_2 < \dfrac{k - I_1}{n-1} \\ 0 & \text{其他} \end{cases}$$

$$\cdots$$

$$I_{l+1} = \begin{cases} 1 & U_{l+1} < \dfrac{k - (I_1 + \cdots + I_l)}{n-l} \\ 0 & \text{其他} \end{cases}$$

$$\cdots$$

$$I_n = \begin{cases} 1 & U_n < \dfrac{k - (I_1 + \cdots + I_{n-1})}{n-1} \\ 0 & \text{其他} \end{cases}$$

如此产生的 $(I_1, \cdots, I_n)$, 其条件分布就是所要求的条件分布, 且其联合分布满足

$$P\{(I_1, \cdots, I_n) = (i_1, \cdots, i_n)\} = 1 \Big/ \binom{n}{k}$$

注释 上述提供的方法并不需要很多的内存. 10.1 节将介绍一个更快的算法, 但要求更多的内存 (10.1 节介绍的算法利用随机排列的后面 k 个元素). ■

例 2h 令 X, Y, Z 服从 $(0,1)$ 上均匀分布且相互独立, 求 $P\{X \geqslant YZ\}$.

解: 因为

$$f_{X,Y,Z}(x,y,z) = f_X(x)f_Y(y)f_Z(z) = 1 \qquad 0 \leqslant x \leqslant 1, 0 \leqslant y \leqslant 1, 0 \leqslant z \leqslant 1$$

我们有

$$\begin{aligned} P\{X \geqslant YZ\} &= \iiint\limits_{x \geqslant yz} f_{X,Y,Z}(x,y,z)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z = \int_0^1 \int_0^1 \int_{yz}^1 \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z \\ &= \int_0^1 \int_0^1 (1 - yz)\,\mathrm{d}y\,\mathrm{d}z = \int_0^1 \left(1 - \frac{z}{2}\right)\mathrm{d}z = \frac{3}{4} \end{aligned}$$ ■

例 2i (半衰期的概率解释) 令 $N(t)$ 表示某矿物中放射性裂变物质在时刻 t 的含量. 所谓半衰期是一个由经验确定的量. 放射性物质随时间的变化规律如下:

$$N(t) = 2^{-t/h}N(0) \qquad t > 0$$

注意到 $N(h) = N(0)/2$. 其中 h 就称为半衰期. 由 $N(t)$ 的规律可知

$$N(t+s) = 2^{-(s+t)/h}N(0) = 2^{-t/h}N(s)$$

因此, 不管已经消耗了多少时间, 在未来的时间 t 内, 这种物质的裂变规律都是按 $2^{-t/h}$ 的因子减少.

这个规律是通过对包含海量分子的放射性物质的观测得到, 它看来与某种概率解释相吻合. 由于裂变物质的比例不依赖于物质的总量, 也不依赖于裂变时间区段的位置 ($N(t+s)/N(s)$ 与 $N(s)$ 和 s 均无关), 这样可以认为各裂变个体的寿命是相互独立的, 并且是公共分布为无记忆的寿命分布. 而唯一的无记忆的寿命分布为指数分布. 由于裂变物质的半衰期为 h, 我们建立下列概率模型.

半衰期 h 的概率解释: 各放射性裂变分子的寿命分布为相互独立的指数分布, 分布的中位数为 h, 记 L 为裂变分子的寿命, 则

$$P\{L < t\} = 1 - 2^{-t/h}$$

由上式看出 $P\{L < h\} = 1/2$, 同时上式又可写成

$$P\{L < t\} = 1 - \exp\left\{-t \cdot \frac{\ln 2}{h}\right\}$$

这说明 L 的分布确是指数分布, 并且中位数为 h.

现在假设在时刻 0 具有 $N(0)$ 个分子, 在时刻 t 具有 $N(t)$ 个分子. 实际上, 每一个分子在时刻 t 时存活的概率为 $p = 2^{-t/h}$, 这样 $N(t)$ 是一个二项随机变量, 参数为 $n = N(0)$, $p = 2^{-t/h}$. 第 8 章将说明, 当考虑含海量分子的放射性物质在一定时间框架下进行衰减时, 半衰期实际上与一个确定的模型相联系. 然而当考虑确定

数量的放射性元素时, 确定模型的解释与概率模型解释的区别就很明显了.①

现在我们来讨论质子的蜕变问题. 关于质子的蜕变问题是有所争论的. 根据一种理论预测质子的半衰期为 $h = 10^{30}$ 年. 为了对此理论进行检验, 有人建议观察大量的质子, 看它们在一两年内有没有蜕变.(显然不可能观察 10^{30} 年之后再看它们是否有一半蜕变,) 现在假定观察 $N(0) = 10^{30}$ 个质子, 观察期为 C 年. 利用确定模型得到

$$\begin{aligned} N(0) - N(c) &= h(1 - 2^{-c/h}) = \frac{1 - 2^{-c/h}}{1/h} \\ &\approx \lim_{x \to 0} \frac{1 - 2^{-cx}}{x} \quad \text{由于} \frac{1}{h} = 10^{-30} \approx 0 \\ &= \lim_{x \to 0} (c2^{-cx} \ln 2) \quad \text{由洛必达法则} \\ &= c \ln 2 \approx 0.6931c \end{aligned}$$

例如, 在 2 年中预计将会有 1.3863 个质子蜕变. 但是如果在 2 年内没有观察到一个质子蜕变, 那对于原来的假设 (10^{30} 年中会有一半的质子蜕变) 是一个极大的打击.

现在我们用概率模型来比较这一结论, 我们观察 h 个质子, 共观察 C 年. 由于 h 相当大 (10^{30}), 每个质子在 C 年内蜕变的概率非常小. 这样, 在 C 年内蜕变的质子数具有泊松分布, 其参数 $n_p = h \cdot (1 - 2^{-c/h}) \approx 1/2^c$. 这样

$$P\{0\text{个蜕变}\} = \mathrm{e}^{-c \ln 2} = \mathrm{e}^{-\ln(2^c)} = 1/2^c$$

一般情况下,

$$P\{n\text{个蜕变}\} = \frac{2^{-c}[c \ln 2]^n}{n!} \quad n \geqslant 0$$

这样, 尽管在 2 年内蜕变的质子平均数为 1.3863(由确定性模型得到), 但利用随机模型得到 2 年内没有观察到质子蜕变的概率为 1/4. 这个结果不能推翻原来的关于质子的半衰期为 10^{30} 年的假设. ■

注释 独立性是一个对称关系.X 和 Y 相互独立是指它们的联合密度 (在离散情况下的联合分布列) 是它们各自密度 (概率) 的乘积. 因此, X 独立于 Y 与 Y 独立于 X 或者 X、Y 相互独立是完全相同的意思. 我们在考虑 X 是否独立于 Y 的时候, 必须考察 Y 的值的改变, 是否会改变 X 的条件分布. 有时候, 我们不能很直观地作出判断. 但是若把 X 和 Y 的地位对调过来, 考虑 Y 是否独立于 X, 此时问题变得十分明显. 下面的例子就说明了这一道理.

例 2j 名为 "Craps" 的掷骰子游戏规定每次抛掷两颗骰子, 若第一次两颗骰子的点数之和为 4, 此时玩家必须继续抛掷这两颗骰子, 直到出现点数之和为 4 或

---

① 在第 8 章中, 我们将介绍大数律. 大数律正好说明 $\frac{N(t)}{N(0)} \sim p = 2^{-t/2}$. 由于 $N(0)$ 和 $N(t)$ 都是很大, 可以认为 $N(t) = 2^{-t/h}N(0)$. —— 译者注

7 为止. 如果最后掷出的点数和为 4, 那么玩家赢, 如果为 7, 则玩家输. 为了更好地了解游戏中某些量之间的关系, 我们将这个游戏简化为下列掷骰子试验. 每次掷两颗骰子, 直到出现两颗骰子的点数之和为 4 或 7 为止. 令 N 表示试验停止的时候, 所投掷的次数. 令 X 表示最后一次投掷的点数之和 (或者 4 或者 7). 那么 N 是否与 X 独立? 也即, 知道了 X 的值, 会不会影响 N 的分布? 很多人并不能从直觉找到这个问题的答案. 然而, 如果我们换个角度问 X 是否独立于 N, 也即, 知道了头一次出现点数之和为 4 或 7 时的抛掷次数, 会不会影响到该点数之和究竟是 4 (或是 7) 的概率? 举例来说, 假设我们已经掷了 $n-1$ 次骰子, 各次均未出现点数之和为 4 或 7, 但第 n 次出现了 4 或 7. 现在问 n 的值会不会影响到 $X=4$ 或 $X=7$ 的概率. 当然不会, 因为所关心的为 4 或 7, 而前 $n-1$ 次抛掷既没出现 4 也没出现 7 这一事实不会改变第 n 次抛掷的概率. 因此, 我们可得出 X 独立于 N, 或者等价地, N 独立于 X 的结论.

另一个例子, 令 $X_1, X_2, \cdots$ 为一系列独立同分布随机变量, 假设我们按顺序观测这些随机变量, 如果 $X_n > X_i$, 对任意 $i=1,\cdots,n-1$ 成立, 那么我们称 X_n 为**记录值**(record value). 也即, 在序列中任何一个值, 若它比其前面的值都大则称这个值为一个记录值. 令 A_n 表示事件"X_n 为一记录值", 那么 A_{n+1} 是否独立于 A_n? 也即, 知道了第 n 个随机变量为前 n 个随机变量的最大值是否会改变第 $n+1$ 个随机变量为前 $n+1$ 个随机变量最大值的概率? 尽管 A_{n+1} 与 A_n 独立, 但是却不是凭直观显而易见的. 然而, 如果我们换个方式, 问 A_n 是否独立于 A_{n+1}? 那么结论是非常容易理解的. 因为知道了第 $n+1$ 个随机变量比 $X_1,\cdots,X_n$ 都大, 这个事实显然没有提供任何关于 X_n 在前 n 个随机变量之间的顺序的信息. 事实上, 利用对称性可得, 这 n 个随机变量里, 任何一个为最大值的可能性都是一样的, 因此 $P(A_n|A_{n+1}) = P(A_n) = 1/n$. 总而言之, A_n 和 A_{n+1} 是相互独立事件. ■

注释 从下面的等式

$$
\begin{aligned}
&P\{X_1 \leqslant a_1, \cdots, X_n \leqslant a_n\} \\
=&P\{X_1 \leqslant a_1\}P\{X_2 \leqslant a_2|X_1 \leqslant a_1\}\cdots P\{X_n \leqslant a_n|X_1 \leqslant a_1, \cdots, X_{n-1} \leqslant a_{n-1}\}
\end{aligned}
$$

可知, 要证明 $X_1,\cdots,X_n$ 的相互独立性, 可以通过序贯的方式加以证明. 也即, 我们要证明这些随机变量是独立的, 只需证明以下的一系列事实

$$
\begin{aligned}
&X_2 \text{ 独立于 } X_1 \\
&X_3 \text{ 独立于 } X_1, X_2 \\
&X_4 \text{ 独立于 } X_1, X_2, X_3 \\
&\cdots \\
&X_n \text{ 独立于 } X_1, \cdots, X_{n-1}
\end{aligned}
$$

6.3 独立随机变量的和

当随机变量 X 和 Y 独立时, 从它们的联合分布求出 $X+Y$ 的分布常常是很重要的. 假设 X 和 Y 是相互独立的连续型随机变量, 其密度函数分别为 f_X 和 f_Y, 那么 $X+Y$ 的分布函数可以如下得到:

$$\begin{aligned}F_{X+Y}(a) &= P\{X+Y \leqslant a\} = \iint\limits_{x+y\leqslant a} f_X(x)f_Y(y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{a-y} f_X(x)f_Y(y)\,\mathrm{d}x\,\mathrm{d}y = \int_{-\infty}^{\infty}\int_{-\infty}^{a-y} f_X(x)\,\mathrm{d}x\, f_Y(y)\,\mathrm{d}y \\ &= \int_{-\infty}^{\infty} F_X(a-y)f_Y(y)\,\mathrm{d}y \end{aligned} \tag{3.1}$$

分布函数 F_{X+Y} 称为分布函数 F_X 和 F_Y (分别表示 X 和 Y 的分布函数) 的卷积 (convolution). 通过对 (3.1) 式求导, 我们可得 $X+Y$ 的密度函数 f_{X+Y} 如下:

$$\begin{aligned}f_{X+Y}(a) &= \frac{\mathrm{d}}{\mathrm{d}a}\int_{-\infty}^{\infty} F_X(a-y)f_Y(y)\,\mathrm{d}y = \int_{-\infty}^{\infty}\frac{\mathrm{d}}{\mathrm{d}a}F_X(a-y)f_Y(y)\,\mathrm{d}y \\ &= \int_{-\infty}^{\infty} f_X(a-y)f_Y(y)\,\mathrm{d}y\end{aligned} \tag{3.2}$$

6.3.1 均匀分布的随机变量[①]

对于两个相互独立的, $(0,1)$ 上均匀分布的随机变量, 其和的分布是不难求的.

例 3a (两个独立均匀分布随机变量的和) 设 X 和 Y 为独立随机变量, 都服从 $(0,1)$ 上均匀分布. 求 $X+Y$ 的概率密度.

解: 因为

$$f_X(a) = f_Y(a) = \begin{cases} 1 & 0 < a < 1 \\ 0 & \text{其他} \end{cases}$$

利用 (3.2), 我们可以得到

$$f_{X+Y}(a) = \int_0^1 f_X(a-y)\,\mathrm{d}y$$

先设 $0 \leqslant a \leqslant 1$, 于是[②]

$$f_{X+Y}(a) = \int_0^a \mathrm{d}y = a$$

当 $1 < a < 2$ 时, 类似地可得

$$f_{X+Y}(a) = \int_{a-1}^1 \mathrm{d}y = 2-a$$

① 本节作者只讨论了 $(0,1)$ 上的均匀分布的随机变量, 但不难推广到一般情况. —— 译者注

② 这是因为当 $y > a$ 时, $f_X(a-y) = 0$; 当 $0 < y < a$ 时, $f_X(a-y) = 1$. —— 译者注

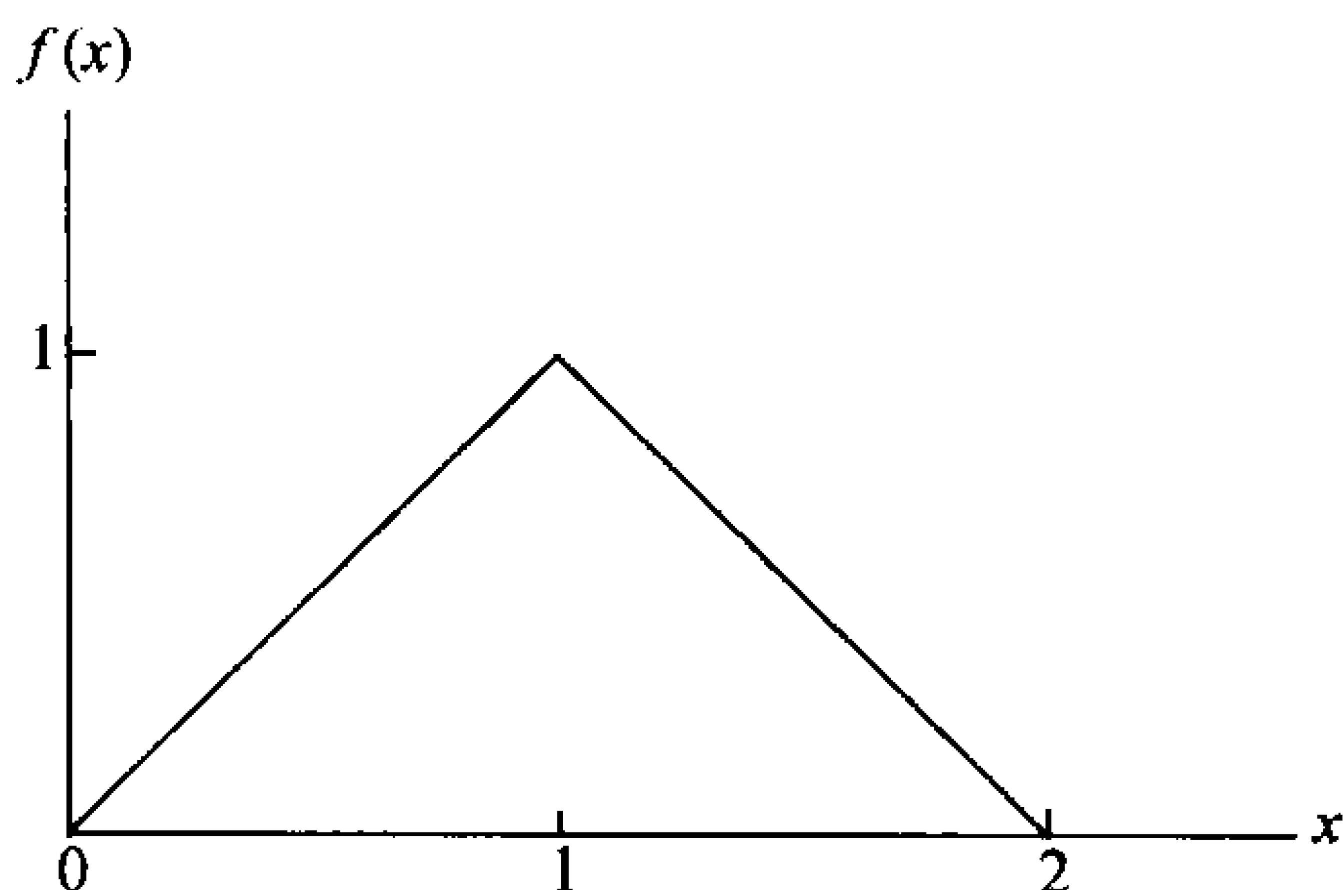


图 6.3　三角分布

因此,

$$f_{X+Y}(a)=\begin{cases}a & 0\leqslant a\leqslant 1\\ 2-a & 1<a<2\\ 0 & \text{其他}\end{cases}$$

$X+Y$ 的密度函数的形状见图 6.3, 其密度函数像一个搁在 x 轴上的三角形, 因此 $X+Y$ 的分布又称为三角分布(triangular). ■

现在假定 $X_1,X_2,\cdots,X_n$ 为独立同分布的 $(0,1)$ 均匀随机变量序列. 记

$$F_n(x)=P\{X_1+\cdots+X_n\leqslant x\}$$

$F_n(x)$ 的分析表达式十分复杂, 然而, 当 $x\leqslant 1$ 时, $F_n(x)$ 具有很好的表达式. 现在我们用归纳法证明

$$F_n(x)=x^n/n!,\quad 0\leqslant x\leqslant 1$$

由于上式当 $n=1$ 时成立. 现假定对 $n-1$ 上式也成立, 即

$$F_{n-1}(x)=x^{n-1}/(n-1)!,\quad 0\leqslant x\leqslant 1$$

现在将 $\sum_{i=1}^n X_i$ 写成

$$\sum_{i=1}^n X_i=\sum_{i=1}^{n-1}X_i+X_n$$

利用 (3.1) 式可得

$$\begin{aligned}F_n(x)&=\int_0^1 F_{n-1}(x-y)f_{X_n}(y)\mathrm{d}y\\&=\frac{1}{(n-1)!}\int_0^x(x-y)^{n-1}\mathrm{d}y\quad \text{由归纳假设}^{①}\\&=\frac{x^n}{n!},\quad 0\leqslant x\leqslant 1\end{aligned}$$

利用 $F_n(x)$ 的这个表达式可以得到一个十分有趣的结果. 设 $X_1,X_2,\cdots$ 为一列独立同分布的 $(0,1)$ 上的均匀随机变量. 我们需要求出下列随机变量 N 的期望 $E[N]$:

$$N=\min\{n:X_1+\cdots+X_n>1\}$$

注意到 $N>n$ 的充要条件为 $X_1+\cdots+X_n\leqslant 1$, 这样

$$P\{N>n\}=P\{X_1+\cdots+X_n\leqslant 1\}=F_n(1)=1/n!,\quad n>0$$

由于 $P\{N>0\}=1=1/0!$, 我们得到

① 即 $\int_0^1 F_{n-1}(x-y)f_{X_n}(y)\mathrm{d}y=\int_0^1 F_n(x-y)\mathrm{d}y=\int_0^x F_{n-1}(x-y)\mathrm{d}y=\frac{1}{(n-1)!}\int_0^x(x-y)^{n-1}\mathrm{d}y.$

—— 译者注

$$P\{N=n\}=P\{N>n-1\}-P\{N>n\}=\frac{1}{(n-1)!}-\frac{1}{n!}=\frac{n-1}{n!}$$

这样

$$E[N]=\sum_{n=1}^{\infty}\frac{n(n-1)}{n!}=\sum_{n=2}^{\infty}\frac{1}{(n-2)!}=\mathrm{e}$$

这样, 总和超过 1 的均匀随机数的 (独立同分布的 $(0,1)$ 均匀随机变量) 最小个数的平均值为 e.

6.3.2 Γ 随机变量

回顾 Γ 随机变量具有如下形式的密度函数:

$$f(y)=\frac{\lambda\mathrm{e}^{-\lambda y}(\lambda y)^{t-1}}{\Gamma(t)}\qquad 0<y<\infty$$

这种分布的一个重要性质就是, 对于给定的 λ, 它在卷积意义下是封闭的.

命题 3.1 如果 X 和 Y 为独立 Γ 随机变量, 参数分别为 (s,λ) 和 (t,λ) 那么 $X+Y$ 也为 Γ 随机变量, 参数为 $(s+t,\lambda)$.

证明: 利用公式 (3.2), 可得

$$\begin{aligned}f_{X+Y}(a)&=\frac{1}{\Gamma(s)\Gamma(t)}\int_0^a\lambda\mathrm{e}^{-\lambda(a-y)}[\lambda(a-y)]^{s-1}\lambda\mathrm{e}^{-\lambda y}(\lambda y)^{t-1}\,\mathrm{d}y\\&=K\mathrm{e}^{-\lambda a}\int_0^a(a-y)^{s-1}y^{t-1}\,\mathrm{d}y\\&=K\mathrm{e}^{-\lambda a}a^{s+t-1}\int_0^1(1-x)^{s-1}x^{t-1}\,\mathrm{d}x\qquad 令x=\frac{y}{a}\\&=C\mathrm{e}^{-\lambda a}a^{s+t-1}\end{aligned}$$

其中 C 为一个不依赖于 a 的常数, 但因为上式为一概率密度函数, 因此积分值等于 1, 这样就确定了 C 的取值. 我们得到

$$f_{X+Y}(a)=\frac{\lambda\mathrm{e}^{-\lambda a}(\lambda a)^{s+t-1}}{\Gamma(s+t)}$$

命题得证. □

利用命题 3.1 及归纳法, 很简单地可以得到: 如果 $X_i, i=1,\cdots,n$ 为独立 Γ 随机变量, 其参数分别为 $(t_i,\lambda), i=1,\cdots,n$, 那么 $\sum_{i=1}^n X_i$ 为参数 $\left(\sum_{i=1}^n t_i,\lambda\right)$ 的 Γ 随机变量, 我们将其证明留作习题.

例 3b 令 $X_1,X_2,\cdots,X_n$ 为 n 个独立指数随机变量, 其公共参数为 λ. 由于参数为 λ 的指数随机变量同时也是参数为 $(1,\lambda)$ 的 Γ 随机变量, 这样通过命题 3.1, 可得 $X_1+X_2+\cdots+X_n$ 为一参数为 (n,λ) 的 Γ 随机变量. ■

如果 $Z_1,Z_2,\cdots,Z_n$ 为相互独立的标准正态随机变量, 那么 $Y\equiv\sum_{i=1}^n Z_i^2$ 称为

服从自由度为 n 的 χ^2 分布的随机变量. 我们来计算其密度函数. 当 $n=1, Y=Z^2$, 利用第 5 章例 7b 可看出, 其密度函数如下:

$$f_{Z^2}(y)=\frac{1}{2\sqrt{y}}[f_Z(\sqrt{y})+f_Z(-\sqrt{y})]=\frac{1}{2\sqrt{y}}\frac{2}{\sqrt{2\pi}}\mathrm{e}^{-y/2}=\frac{\frac{1}{2}\mathrm{e}^{-y/2}(y/2)^{1/2-1}}{\sqrt{\pi}}$$

由上述形式可以看出, $f_{Z^2}(y)$ 为参数 $(1/2,1/2)$ 的 Γ 分布. [从以上分析可以顺便得到结果: $\Gamma(1/2)=\sqrt{\pi}$]. 但由于每个 Z_i^2 都服从 $\Gamma(1/2,1/2)$, 利用命题 3.1 我们可得: 自由度为 n 的 χ^2 分布就是参数为 $(n/2,1/2)$ 的 Γ 分布. 因此, 其密度函数为:

$$f_{\chi^2}(y)=\frac{\frac{1}{2}\mathrm{e}^{-y/2}\left(\frac{y}{2}\right)^{n/2-1}}{\Gamma\left(\frac{n}{2}\right)}=\frac{\mathrm{e}^{-y/2}y^{n/2-1}}{2^{n/2}\Gamma\left(\frac{n}{2}\right)}\qquad y>0$$

当 n 为偶数时, $\Gamma(n/2)=[(n/2)-1]!$, 而 n 为奇数时, $\Gamma(n/2)$ 可以反复利用关系式 $\Gamma(t)=(t-1)\Gamma(t-1)$ 进行计算, 再利用之前得到的结果 $\Gamma(1/2)=\sqrt{\pi}$ 可以得到 $\Gamma(n/2)$ 的值. $\left[\text{举例来说, }\Gamma(5/2)=\frac{3}{2}\Gamma\left(\frac{3}{2}\right)=\frac{3}{2}\frac{1}{2}\Gamma\left(\frac{1}{2}\right)=\frac{3}{4}\sqrt{\pi}.\right]$

χ^2 分布在实践中经常作为以下问题中的误差的平方和的分布出现: 如果某人试图击中 n 维空间里的目标, 其中每个方向上的误差为独立标准正态随机变量, 则各个方向上的误差的平方和的分布为自由度 n 的 χ^2 分布. 它在统计分析中也是很重要的.

6.3.3 正态随机变量

我们还可以利用公式 (3.2) 证明关于正态随机变量的以下的重要结论.

> **命题 3.2** 若 $X_i, i=1,\cdots,n$ 为独立随机变量, 它们均服从正态分布, 各自参数分别为 $(\mu_i,\sigma_i^2), i=1,\cdots,n$, 则 $\sum_{i=1}^n X_i$ 也服从正态分布, 参数为 $\sum_{i=1}^n \mu_i$ 和 $\sum_{i=1}^n \sigma_i^2$.

证明: 首先, 令 X 和 Y 为独立正态随机变量, X 的均值为 0, 方差为 σ^2, 而 Y 的均值为 0, 方差为 1. 我们要利用公式 (3.2) 来计算 $X+Y$ 的密度函数. 设

$$c=\frac{1}{2\sigma^2}+\frac{1}{2}=\frac{1+\sigma^2}{2\sigma^2}$$

这样, 有

$$\begin{aligned}f_X(a-y)f_Y(y)&=\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(a-y)^2}{2\sigma^2}\right\}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{y^2}{2}\right\}\\&=\frac{1}{2\pi\sigma}\exp\left\{-\frac{a^2}{2\sigma^2}\right\}\exp\left\{-c\left(y^2-2y\frac{a}{1+\sigma^2}\right)\right\}\end{aligned}$$

因此, 由公式 (3.2), 有

$$f_{X+Y}(a) = \frac{1}{2\pi\sigma}\exp\left\{-\frac{a^2}{2\sigma^2}\right\}\exp\left\{\frac{a^2}{2\sigma^2(1+\sigma^2)}\right\}\times\int_{-\infty}^{\infty}\exp\left\{-c\left(y-\frac{a}{1+\sigma^2}\right)^2\right\}\mathrm{d}y$$
$$= \frac{1}{2\pi\sigma}\exp\left\{-\frac{a^2}{2(1+\sigma^2)}\right\}\int_{-\infty}^{\infty}\exp\left\{-cx^2\right\}\mathrm{d}x = C\exp\left\{-\frac{a^2}{2(1+\sigma^2)}\right\}$$

其中 C 的取值不依赖于 a. 上式意味着 $X+Y$ 服从均值为 0, 方差为 $1+\sigma^2$ 的正态分布.

现在, 假设 X_1 和 X_2 为独立正态分布随机变量, X_i 的均值为 μ_i, 方差为 σ_i^2, $i=1,2$. 那么,

$$X_1+X_2 = \sigma_2\left(\frac{X_1-\mu_1}{\sigma_2}+\frac{X_2-\mu_2}{\sigma_2}\right)+\mu_1+\mu_2$$

但因为 $(X_1-\mu_1)/\sigma_2$ 服从均值为 0, 方差为 σ_1^2/σ_2^2 的正态分布, 而 $(X_2-\mu)/\sigma_2$ 服从均值为 0, 方差为 1 的正态分布, 利用前面的结果可得 $(X_1-\mu)/\sigma_1+(X_2-\mu)/\sigma_2$ 服从均值为 0, 方差为 $1+\sigma_1^2/\sigma_2^2$ 的正态分布, 这意味着 X_1+X_2 服从均值为 $\mu_1+\mu_2$, 方差为 $\sigma_2^2(1+\sigma_1^2/\sigma_2^2)=\sigma_1^2+\sigma_2^2$ 的正态分布.

因此, 当 $n=2$ 时, 命题 3.2 成立. 一般情形可通过如下归纳法得到, 也即, 假定对于 $n-1$ 个随机变量时成立, 现在考虑 n 的情形,

$$\sum_{i=1}^{n} X_i = \sum_{i=1}^{n-1} X_i + X_n$$

利用归纳假设, $\sum_{i=1}^{n-1} X_i$ 服从均值为 $\sum_{i=1}^{n-1}\mu_i$, 方差为 $\sum_{i=1}^{n-1}\sigma_i^2$ 的正态分布. 因此, 利用 $n=2$ 的结果, 可得 $\sum_{i=1}^{n} X_i$ 服从正态分布, 参数为 $\sum_{i=1}^{n}\mu_i$ 和 $\sum_{i=1}^{n}\sigma_i^2$. □

例 3c 某俱乐部篮球队一个赛季打 44 场比赛. 其中有 26 场是对甲级队, 18 场对乙级队. 假设对甲级队每场胜的概率为 0.4, 而对乙级队每场胜的概率为 0.7. 假设每场比赛结果都是独立的. 近似计算以下事件概率:

(a) 该队能赢 25 场以上比赛;

(b) 该队胜甲级队的场数超过胜乙级队的场数.

解: (a) 令 X_A 和 X_B 分别表示该队同甲级队和乙级队比赛获胜场数, 注意到 X_A 和 X_B 为独立二项随机变量, 及

$$E[X_A] = 26\times 0.4 = 10.4 \quad \mathrm{Var}(X_A) = 26\times 0.4\times 0.6 = 6.24$$
$$E[X_B] = 18\times 0.7 = 12.6 \quad \mathrm{Var}(X_B) = 18\times 0.7\times 0.3 = 3.78$$

利用二项分布的正态近似可得, X_A 和 X_B 都近似服从均值和方差如上的独立正态分布. 因此, 由命题 3.2, X_A+X_B 近似服从均值为 23, 方差为 10.02 的正态分布. 令 Z 表示标准正态随机变量, 我们有

$$P\{X_A+X_B\geqslant 25\}=P\{X_A+X_B\geqslant 24.5\}=P\left\{\frac{X_A+X_B-23}{\sqrt{10.02}}\geqslant\frac{24.5-23}{\sqrt{10.02}}\right\}$$
$$\approx P\left\{Z\geqslant\frac{1.5}{\sqrt{10.02}}\right\}\approx 1-P\{Z<0.4739\}\approx 0.3178$$

(b) 注意到 X_A-X_B 近似服从均值为 -2.2, 方差为 10.02 的正态分布. 因此

$$P\{X_A-X_B\geqslant 1\}=P\{X_A-X_B\geqslant 0.5\}=P\left\{\frac{X_A-X_B+2.2}{\sqrt{10.02}}\geqslant\frac{0.5+2.2}{\sqrt{10.02}}\right\}$$
$$\approx P\left\{Z\geqslant\frac{2.7}{\sqrt{10.02}}\right\}\approx 1-P\{Z<0.8530\}\approx 0.1968$$

因此, 该队近似以 31.87% 的概率赢得 25 场比赛, 与甲级队比赛获胜场数超过与乙级队获胜场数的概率近似为 19.68%. ■

如果 $\ln(Y)$ 为正态随机变量, 均值为 μ, 方差为 σ^2, 那么 Y 称为参数为 (μ,σ) 的**对数正态** (lognormal) **随机变量**. 也即, 如果 Y 能够表示为 $Y=\mathrm{e}^X$, 其中 X 为一正态随机变量, 那么 Y 为对数正态随机变量.

例 3d 令 $S(n)$ 表示从某时刻开始, $n\,(n\geqslant 1)$ 周后某证券的价格. 一个比较流行的关于证券价格运行的模型认为股价比例 $S(n)/S(n-1), n\geqslant 1$ 为独立同分布对数正态随机变量, 假设该模型的参数为, $\mu=0.0165,\sigma=0.0730$, 试求以下事件概率:

(a) 接下来的两周证券价格都在上涨;

(b) 两周后的证券价格比现在的高.

解: 令 Z 为一标准正态随机变量. 为了解答 (a), 我们利用以下事实, $\ln(x)$ 对于 x 递增蕴含着 "$x>1$ 当且仅当 $\ln(x)>\ln(1)=0$". 这样, 我们有

$$P\left\{\frac{S(1)}{S(0)}>1\right\}=P\left\{\ln\left(\frac{S(1)}{S(0)}\right)>0\right\}$$
$$=P\left\{Z>\frac{-0.0165}{0.0730}\right\}=P\{Z<0.2260\}=0.5894$$

因此, 一周后价格上涨的概率为 0.5894. 由于连续的价格比独立, 因此, 接下来的两周价格都上涨的概率为 $0.5894^2=0.3474$.

(b) 的解答如下:

$$P\left\{\frac{S(2)}{S(0)}>1\right\}=P\left\{\frac{S(2)S(1)}{S(1)S(0)}>1\right\}=P\left\{\ln\left(\frac{S(2)}{S(1)}\right)+\ln\left(\frac{S(1)}{S(0)}\right)>0\right\}$$

由于 $\ln\left(S(2)/S(1)\right)$ 和 $\ln\left(S(1)/S(0)\right)$ 为独立同分布的正态随机变量, 公共参数为 $\mu_i=0.0165,\sigma^2=0.0730^2$, 它们和的期望为 $2\mu_i=0.0330$, 方差为 2×0.0730^2, 因此

$$P\left\{\frac{S(2)}{S(0)}>1\right\}=P\left\{Z>\frac{-0.0330}{0.0730\sqrt{2}}\right\}=P\{Z<0.31965\}=0.6254$$ ■

6.3.4 泊松随机变量和二项随机变量

与其试图推导出离散型情形下的 $X+Y$ 的分布函数的一般表达式, 不如来看看以下几个例子:

例 3e (独立泊松分布随机变量的和) 设 X 和 Y 为独立泊松分布随机变量, 参数分别为 λ_1 和 λ_2, 求 $X+Y$ 的分布.

解: 因为事件 $\{X+Y=n\}$ 能够写成互不相容事件 $\{X=k, Y=n-k\}, 0 \leqslant k \leqslant n$ 的并, 因此有

$$
\begin{aligned}
P\{X+Y=n\} &= \sum_{k=0}^{n} P\{X=k, Y=n-k\} = \sum_{k=0}^{n} P\{X=k\}P\{Y=n-k\} \\
&= \sum_{k=0}^{n} \mathrm{e}^{-\lambda_1}\frac{\lambda_1^k}{k!}\mathrm{e}^{-\lambda_2}\frac{\lambda_2^{n-k}}{(n-k)!} = \mathrm{e}^{-(\lambda_1+\lambda_2)}\sum_{k=0}^{n}\frac{\lambda_1^k\lambda_2^{n-k}}{k!(n-k)!} \\
&= \frac{\mathrm{e}^{-(\lambda_1+\lambda_2)}}{n!}\sum_{k=0}^{n}\frac{n!}{k!(n-k)!}\lambda_1^k\lambda_2^{n-k} = \frac{\mathrm{e}^{-(\lambda_1+\lambda_2)}}{n!}(\lambda_1+\lambda_2)^n
\end{aligned}
$$

也就是说, X_1+X_2 服从参数为 $\lambda_1+\lambda_2$ 的泊松分布. ■

例 3f (独立二项分布随机变量的和) 令 X 和 Y 为独立的二项分布随机变量, 参数分别为 (n,p) 和 (m,p), 求 $X+Y$ 的分布.

解: 回顾二项分布的知识, 根本不用任何计算, 我们也可以马上推导出, $X+Y$ 服从参数为 $(n+m,p)$ 的二项分布. 这是因为 X 表示了 n 次独立重复试验中成功的次数 (每次成功的概率为 p), 同样, Y 表示了 m 次独立重复试验中成功的次数 (每次成功的概率也为 p), 由于假定 X 和 Y 独立, 所以 $X+Y$ 表示了在 $n+m$ 次独立重复试验中成功的次数 (每次成功的概率为 p). 这样, $X+Y$ 为参数为 $(n+m,p)$ 的二项随机变量. 下面从分析的角度验证该结论, 注意到

$$
\begin{aligned}
P\{X+Y=k\} &= \sum_{i=0}^{n} P\{X=i, Y=k-i\} \\
&= \sum_{i=0}^{n} P\{X=i\}P\{Y=k-i\} = \sum_{i=0}^{n}\binom{n}{i}p^i q^{n-i}\binom{m}{k-i}p^{k-i}q^{m-k+i}
\end{aligned}
$$

其中 $q=1-p$ 且当 $j>r$ 时 $\binom{r}{j}=0$. 因此

$$
P\{X+Y=k\} = p^k q^{n+m-k}\sum_{i=0}^{n}\binom{n}{i}\binom{m}{k-i}
$$

最后利用以下组合恒等式便可得到所需结果:

$$
\binom{n+m}{k} = \sum_{i=0}^{n}\binom{n}{i}\binom{m}{k-i}
$$

■

6.3.5 几何随机变量

设 $X_1, \cdots, X_n$ 是独立的几何随机变量序列, X_i 的分布参数为 $p_i, i=1,2,\cdots,n$. 现在我们要求出和 $S_n=\sum_{i=1}^n X_i$ 的分布列. 作为应用, 考虑 n 个硬币, 假定第 i 个硬币在抛掷时正面向上的概率为 $p_i, i=1,\cdots,n$. 假定将第 1 枚硬币进行抛掷, 直到出现正面向上为止, 同样, 将第 2 枚硬币也抛掷到出现正面朝上为止, 等等. 令 X_i 为第 i 枚硬币的抛掷次数, 则 $X_1,\cdots,X_n$ 是独立的几何随机变量序列, 其分布参数分别为 $p_1,p_2,\cdots,p_n$, 此时 S_n 就是抛掷硬币的总次数. 若所有 $p_i=p$, 此时 S_n 就是连续抛掷硬币, 直到出现第 n 次正面向上时的抛掷总次数. S_n 具有负二项分布, 其分布列为

$$P\{S_n=k\}=\binom{k-1}{n-1}p^n(1-p)^{k-n}, \quad k \geqslant n$$

上述公式只是我们所解决问题的一个前凑. 现在解决所有 p_i 不同的情况. 先解决 $n=2$ 的情况, 记 $q_j=1-p_j, j=1,2$, 我们有

$$\begin{aligned}
P(S_2=k) &= \sum_{j=1}^{k-1} P\{X_1=j, X_2=k-j\} \\
&= \sum_{j=1}^{k-1} P\{X_1=j\}P\{X_2=k-j\} \quad (\text{由于 } X_1, X_2 \text{ 相互独立}) \\
&= \sum_{j=1}^{k-1} p_1 q_1^{j-1} p_2 q_2^{k-j-1} = p_1 p_2 q_2^{k-2} \sum_{j=1}^{k-1} (q_1/q_2)^{j-1} = p_1 p_2 q_2^{k-2} \frac{1-(q_1/q_2)^{k-1}}{1-q_1/q_2} \\
&= \frac{p_1 p_2 q_2^{k-1}}{q_2-q_1} - \frac{p_1 p_2 q_1^{k-1}}{q_2-q_1} = p_2 q_2^{k-1}\frac{p_1}{p_1-p_2} + p_1 q_1^{k-1}\frac{p_2}{p_2-p_1}
\end{aligned}$$

现在令 $n=3$, 并计算 $P\{S_3=k\}$, 将已得到的关于 $P\{S_2=j\}$ 的公式代入 $P\{S_3=k\}$ 的下面的表达式

$$P\{S_3=k\}=\sum_{j=1}^{k-1} P\{S_2=j, X_3=k-j\}=\sum_{j=1}^{k-1} P\{S_2=j\}P\{X_3=k-j\}$$

经过一些运算, 最后得到

$$\begin{aligned}
P\{S_3=k\} =& p_1 q_1^{k-1}\frac{p_2}{p_2-p_1}\cdot\frac{p_3}{p_3-p_1} + p_2 q_2^{k-1}\frac{p_1}{p_1-p_2}\cdot\frac{p_3}{p_3-p_2} \\
&+ p_3 q_3^{k-1}\frac{p_1}{p_1-p_3}\cdot\frac{p_2}{p_2-p_3}
\end{aligned}$$

这个公式很有规则, 由上式以及 $P\{S_2=k\}$ 的公式, 可以猜测到 $P\{S_n=k\}$ 的一般公式. 下面是关于 $P\{S_n=k\}$ 的一个命题.

命题 3.3 设 $X_1, \cdots, X_n$ 为一独立的几何随机变量序列, X_i 的分布参数为 p_i, 假定 p_i 都不相同, 则对于 $k \geqslant n$,

$$P\{S_n = k\} = \sum_{i=1}^{n} p_i q_i^{k-1} \prod_{j \neq i} \frac{p_j}{p_j - p_i}$$

证明: 我们用对 $n+k$ 作归纳法证明此命题. 对于 $n=2, k=2$, 命题是正确的. 现在假定, 对一切 $k \geqslant n, n+k \leqslant r$ 命题成立. 现在令 $k \geqslant n$, 并且 $n+k=r+1$. 利用全概公式

$$\begin{aligned} P\{S_n = k\} &= P\{S_n = k | X_n = 1\} P\{X_n = 1\} + P\{S_n = k | X_n > 1\} P\{X_n > 1\} \\ &= P\{S_n = k | X_n = 1\} p_n + P\{S_n = k | X_n > 1\} q_n \end{aligned}$$

对于条件 $\{X_n = 1\}$, 我们有

$$\begin{aligned} P\{S_n = k | X_n = 1\} &= P\{S_{n-1} = k-1 | X_n = 1\} \\ &= P\{S_{n-1} = k-1\} \quad (\text{由 } X_n \text{ 与 } S_{n-1} \text{ 相互独立}) \\ &= \sum_{i=1}^{n-1} p_i q_i^{k-2} \prod_{i \neq j \leqslant n-1} \frac{p_j}{p_j - p_i} \quad (\text{由归纳法假设}) \end{aligned}$$

现在设 X 是一个参数为 p 的几何随机变量, 经过计算可知, 在 $X>1$ 的条件下 X 的分布与 $1+Y$ 的分布相同, 其中 Y 的分布与 X 的分布是相同的. 同样, S_n 在 $X_n > 1$ 下条件分布与 $X_1 + \cdots + X_{n-1} + Y + 1$ 的分布相同, 其中 Y 与 X_n 的分布相同, 并且 Y 与 $X_1, \cdots, X_{n-1}$ 是相互独立的. 这样, 我们得到

$$\begin{aligned} P\{S_n = k | X_n > 1\} &= P\{X_1 + \cdots + X_{n-1} + X_n + 1 = k\} \\ &= P\{S_n = k-1\} \\ &= \sum_{i=1}^{n} p_i q_i^{k-2} \prod_{i \neq j \leqslant n} \frac{p_j}{p_j - p_i} \quad (\text{由归纳法假设}) \end{aligned}$$

在得到 $P\{S_n = k | X_n = 1\}$ 与 $P\{S_n = k | X_n > 1\}$ 的表达式以后, 由全概公式可得

$$\begin{aligned} P\{S_n = k\} =& p_n \sum_{i=1}^{n-1} p_i q_i^{k-2} \prod_{i \neq j \leqslant n-1} \frac{p_j}{p_j - p_i} + q_n \sum_{i=1}^{n} p_i q_i^{k-2} \prod_{i \neq j \leqslant n} \frac{p_j}{p_j - p_i} \\ =& p_n \sum_{i=1}^{n-1} p_i q_i^{k-2} \prod_{i \neq j \leqslant n-1} \frac{p_j}{p_j - p_i} + q_n \sum_{i=1}^{n-1} p_i q_i^{k-2} \prod_{i \neq j \leqslant n} \frac{p_j}{p_j - p_i} \\ &+ q_n p_n q_n^{k-2} \prod_{j<n} \frac{p_j}{p_j - p_n} \\ =& \sum_{i=1}^{n-1} p_i q_i^{k-2} p_n \left(1 + \frac{q_n}{p_n - p_i}\right) \prod_{i \neq j \leqslant n-1} \frac{p_j}{p_j - p_i} + p_n q_n^{k-1} \prod_{j<n} \frac{p_j}{p_j - p_n} \end{aligned}$$

现在利用恒等式

$$1+\frac{q_n}{p_n-p_i}=\frac{p_n-p_i+q_n}{p_n-p_i}=\frac{q_i}{p_n-p_i}$$

前式变成

$$\begin{aligned}P\{S_n=k\}&=\sum_{i=1}^{n-1}p_iq_i^{k-1}\prod_{i\neq j\leqslant n}\frac{p_j}{p_j-p_i}+p_nq_n^{k-1}\prod_{j<n}\frac{p_j}{p_j-p_n}\\&=\sum_{i=1}^{n}p_iq_i^{k-1}\prod_{j\neq i}\frac{p_j}{p_j-p_i}\end{aligned}$$

由归纳法知, 命题得证. □

6.4 离散情形下的条件分布

回顾对于任两个事件 E 和 F, 已知 F 的条件下 E 的条件概率 (假定 $P(F)>0$) 定义如下:

$$P(E|F)=\frac{P(EF)}{P(F)}$$

因此, 如果 X 和 Y 都是离散型随机变量, 那么很自然地定义已知 $Y=y$ 的条件下 X 的分布列如下: 对于所有满足 $p_Y(y)>0$ 的 y, 有

$$p_{X|Y}(x|y)=P\{X=x|Y=y\}=\frac{P\{X=x,Y=y\}}{P\{Y=y\}}=\frac{p(x,y)}{p_Y(y)}$$

类似地, 也可以定义已知 $Y=y$ 的条件下 X 的条件分布函数, 对于所有满足 $p_Y(y)>0$ 的 y, 有

$$F_{X|Y}(x|y)=P\{X\leqslant x|Y=y\}=\sum_{a\leqslant x}p_{X|Y}(a|y)$$

由上述定义可知, 条件分布与普通分布在概念上是完全一样的, 只是所涉及的事件都是在 $Y=y$ 之条件意义下的事件. 如果 X 和 Y 独立, 那么条件分布列和条件分布函数同通常分布列和分布函数是一样的. 这是因为, 如果 X 和 Y 独立, 那么

$$\begin{aligned}p_{X|Y}(x|y)&=P\{X=x|Y=y\}\\&=\frac{P\{X=x,Y=y\}}{P\{Y=y\}}=\frac{P\{X=x\}P\{Y=y\}}{P\{Y=y\}}=P\{X=x\}\end{aligned}$$

例 4a 假设 X 和 Y 的联合分布列 $p(x,y)$ 如下:

$$p(0,0)=0.4\quad p(0,1)=0.2\quad p(1,0)=0.1\quad p(1,1)=0.3$$

求已知 $Y=1$ 的条件下 X 的条件分布列.

解: 首先注意到

$$p_Y(1)=\sum_x p(x,1)=p(0,1)+p(1,1)=0.5$$

因此,

$$p_{X|Y}(0|1)=\frac{p(0,1)}{p_Y(1)}=\frac{2}{5}\quad 及\quad p_{X|Y}(1|1)=\frac{p(1,1)}{p_Y(1)}=\frac{3}{5}\qquad\blacksquare$$

例 4b 如果 X 和 Y 为独立泊松随机变量, 参数分别为 λ_1 和 λ_2, 求给定 $X+Y=n$ 的条件下 X 的条件分布.

解: 给定 $X+Y=n$ 条件下 X 的条件分布列计算如下:

$$P\{X=k|X+Y=n\}=\frac{P\{X=k,X+Y=n\}}{P\{X+Y=n\}}$$
$$=\frac{P\{X=k,Y=n-k\}}{P\{X+Y=n\}}=\frac{P\{X=k\}P\{Y=n-k\}}{P\{X+Y=n\}}$$

其中最后一个等式的成立是由于假定了 X 和 Y 是独立的. 注意到 (例 3e) $X+Y$ 服从参数为 $\lambda_1+\lambda_2$ 的泊松分布, 这样

$$P\{X=k|X+Y=n\}=\frac{\mathrm{e}^{-\lambda_1}\lambda_1^k}{k!}\frac{\mathrm{e}^{-\lambda_2}\lambda_2^{n-k}}{(n-k)!}\left[\frac{\mathrm{e}^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^n}{n!}\right]^{-1}$$
$$=\frac{n!}{(n-k)!k!}\frac{\lambda_1^k\lambda_2^{n-k}}{(\lambda_1+\lambda_2)^n}=\binom{n}{k}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}$$

也就是说, 给定 $X+Y=n$ 的条件下的 X 的条件分布为二项分布, 参数为 $(n, \lambda_1/(\lambda_1+\lambda_2))$. ■

我们还可以讨论联合条件分布, 正如下面的两个例子所阐述的.

例 4c 考虑如下的多项分布:

$$P\{X_i=n_i,i=1,\cdots,k\}=\frac{n\,!}{n_1\,!\cdots n_k\,!}p_1^{n_1}\cdots p_k^{n_k},\qquad n_i\geqslant 0,\sum_{i=1}^k n_i=n$$

这样的分布列出现在下面的情形: 进行 n 次独立重复试验, 每次试验的第 i 个结果发生的概率为 $p_i,\sum_{i=1}^k p_i=1$. 随机变量 $X_i,\ i=1,\cdots,k$ 分别表示 n 次试验中试验结果 $i,i=1,\cdots,k$ 出现的次数. 现在假设我们已经知道在 n 次试验中, 第 j 个试验结果出现了 n_j 次, $j=r+1,\cdots,k$, 其中 $\sum_{j=r+1}^k n_j=m<n$. 由于另外的 $n-m$ 个试验中每一个结果都必然是第 $1,\cdots,r$ 个结果之一, 因此, $X_1,\cdots,X_r$ 的条件分布应该是多项分布, 其相应的参数为

$$P\{第\ i\ 个结果|试验结果不是\ r+1,\cdots,k\ 中任一结果\}=\frac{p_i}{F_r}\qquad i=1,\cdots,r$$

其中 $F_r=\sum_{i=1}^r p_i$ 为试验结果为 $1,\cdots,r$ 之一的概率.

解: 为了验证该直观结论, 令 $n_1,\cdots,n_r$ 满足 $\sum_{i=1}^r n_i=n-m$, 那么

$$P\{X_1=n_1,\cdots,X_r=n_r|X_{r+1}=n_{r+1},\cdots,X_k=n_k\}$$
$$=\frac{P\{X_1=n_1,\cdots,X_k=n_k\}}{P\{X_{r+1}=n_{r+1},\cdots,X_k=n_k\}}=\frac{\frac{n\,!}{n_1\,!\cdots n_k\,!}p_1^{n_1}\cdots p_r^{n_r}p_{r+1}^{n_{r+1}}\cdots p_k^{n_k}}{\frac{n\,!}{(n-m)\,!\,n_{r+1}\,!\cdots n_k\,!}F_r^{n-m}p_{r+1}^{n_{r+1}}\cdots p_k^{n_k}}$$

上式分母中的概率计算如下, 将试验结果 $1, 2, \cdots, r$ 合成一个结果, 其相应的概率为 F_r, 相应出现的次数为 $n-m$. 这样, 事件 $\{X_{r+1}=n_{r+1}, \cdots, X_k=n_k\}$ 的概率可以看作是 n 次试验中, 具有 $k-r+1$ 个结果的多项分布中事件的概率. 将上式中分子分母相互抵消, 可得

$$
\begin{aligned}
& P\{X_1=n_1, \cdots, X_r=n_r | X_{r+1}=n_{r+1}, \cdots, X_k=n_k\} \\
& = \frac{(n-m)!}{n_1! \cdots n_r!}\left(\frac{p_1}{F_r}\right)^{n_1} \cdots \left(\frac{p_r}{F_r}\right)^{n_r}
\end{aligned}
$$

这样结论得证. ■

例 4d 考虑 n 次独立重复试验, 每次成功概率为 p, 已知共有 k 次成功的条件下, 证明所有可能的 k 次成功、$n-k$ 次失败的顺序都是等可能的.

解: 我们将要证明, $\binom{n}{k}$ 种可能顺序中任一种都是等可能的. 令 X 表示成功的次数, 考虑 k 次成功、$n-k$ 次失败的任一排列, 比如, $\mathbf{o}=(s, s, f, f, \cdots, f)$, 那么

$$
P(\mathbf{o}|X=k) = \frac{P(\mathbf{o}, X=k)}{P(X=k)} = \frac{P(\mathbf{o})}{P(X=k)} = \frac{p^k(1-p)^{n-k}}{\binom{n}{k}p^k(1-p)^{n-k}} = \frac{1}{\binom{n}{k}}
$$
■

6.5 连续情形下的条件分布

如果 X 和 Y 具有联合密度函数 $f(x, y)$, 那么给定条件 $Y=y$ 下, X 的条件密度函数定义如下, 对于任意满足 $f_Y(y)>0$ 的 y 值, 有

$$
f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}
$$

为了说明条件密度的实际含义, 在上式左边乘以 $\mathrm{d}x$, 右边乘以 $(\mathrm{d}x\,\mathrm{d}y)/\mathrm{d}y$, 这样可得

$$
\begin{aligned}
f_{X|Y}(x|y)\mathrm{d}x &= \frac{f(x, y)\,\mathrm{d}x\,\mathrm{d}y}{f_Y(y)\,\mathrm{d}y} \approx \frac{P\{x \leqslant X \leqslant x+\mathrm{d}x, y \leqslant Y \leqslant y+\mathrm{d}y\}}{P\{y \leqslant Y \leqslant y+\mathrm{d}y\}} \\
&= P\{x \leqslant X \leqslant x+\mathrm{d}x \,|\, y \leqslant Y \leqslant y+\mathrm{d}y\}
\end{aligned}
$$

也就是说, 对于很小的 $\mathrm{d}x$ 和 $\mathrm{d}y$, $f_{X|Y}(x|y)\,\mathrm{d}x$ 表示了在 Y 取值于 y 和 $y+\mathrm{d}y$ 之间的条件下, X 取值于 x 和 $x+\mathrm{d}x$ 之间的条件概率.

利用条件密度, 还可以定义已知一个随机变量的取值条件下, 关于另外一个随机变量的事件的条件概率. 也即, 如果 X 和 Y 联合连续, 那么对任一集合 A, 有

$$
P\{X \in A \,|\, Y=y\} = \int_A f_{X|Y}(x|y)\,\mathrm{d}x
$$

特别地, 令 $A=(-\infty, a]$, 那么已知 $Y=y$ 的条件下 X 的条件分布函数如下:

$$F_{X|Y}(a|y) \equiv P\{X \leqslant a \,|\, Y=y\} = \int_{-\infty}^{a} f_{X|Y}(x|y)\,\mathrm{d}x$$

读者应该注意到, 上述讨论已经给出了条件概率的较完整的定义, 即使条件 $Y=y$ 是一个零概率事件, 相应的条件概率也有较明确的含义.

例 5a X 和 Y 的联合密度如下:

$$f(x,y)=\begin{cases}\dfrac{12}{5}x(2-x-y) & 0<x<1, 0<y<1\\ 0 & \text{其他}\end{cases}$$

求已知 $Y=y$ 的条件下, X 的条件密度, 其中 $0<y<1$.

解: 对于 $0<x<1, 0<y<1$, 有

$$\begin{aligned}f_{X|Y}(x|y) &= \frac{f(x,y)}{f_Y(y)} = \frac{f(x,y)}{\int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}x}\\ &= \frac{x(2-x-y)}{\int_0^1 x(2-x-y)\,\mathrm{d}x} = \frac{x(2-x-y)}{\frac{2}{3}-y/2} = \frac{6x(2-x-y)}{4-3y}\end{aligned}$$ ■

例 5b 假设 X 和 Y 的联合密度如下:

$$f(x,y)=\begin{cases}\dfrac{\mathrm{e}^{-x/y}\mathrm{e}^{-y}}{y} & 0<x<\infty, 0<y<\infty\\ 0 & \text{其他}\end{cases}$$

求 $P\{X>1|Y=y\}$.

解: 我们先求得 $Y=y$ 条件下 X 的条件密度:

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{\mathrm{e}^{-x/y}\mathrm{e}^{-y}/y}{\mathrm{e}^{-y}\int_0^{\infty}(1/y)\mathrm{e}^{-x/y}\,\mathrm{d}x} = \frac{1}{y}\mathrm{e}^{-x/y}$$

因此,

$$P\{X>1|Y=y\} = \int_1^{\infty}\frac{1}{y}\mathrm{e}^{-x/y}\,\mathrm{d}x = -\mathrm{e}^{-x/y}\Big|_1^{\infty} = \mathrm{e}^{-1/y}$$ ■

如果 X 和 Y 为独立连续型随机变量, 那么给定 $Y=y$ 的条件下, X 的条件密度同非条件密度是一样的, 这是因为在独立情形下,

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{f_X(x)f_Y(y)}{f_Y(y)} = f_X(x)$$

当随机变量既非联合连续, 也非联合离散, 我们也可考虑相应的条件分布. 举例来说, 假设 X 为一连续型随机变量, 其密度函数为 f, 而 N 为一离散型随机变量, 那么考虑给定 $N=n$ 的条件下 X 的条件分布, 有

$$\frac{P\{x<X<x+\mathrm{d}x|N=n\}}{\mathrm{d}x}$$

$$= \frac{P\{N=n|x<X<x+\mathrm{d}x\}}{P\{N=n\}} \frac{P\{x<X<x+\mathrm{d}x\}}{\mathrm{d}x}$$

令 $\mathrm{d}x$ 趋于 0，可得

$$\lim_{\mathrm{d}x\to 0} \frac{P\{x<X<x+\mathrm{d}x|N=n\}}{\mathrm{d}x} = \frac{P\{N=n|X=x\}}{P\{N=n\}} f(x)$$

这样就证明了给定 $N=n$ 的条件下, X 的条件密度为

$$f_{X|N}(x|n) = \frac{P\{N=n|X=x\}}{P\{N=n\}} f(x)$$

例 5c (二元正态分布) 二元正态分布是非常重要的联合分布之一. 如果对常数 $\mu_x, \mu_y, \sigma_x > 0, \sigma_y > 0, -1 < \rho < 1$, 以及所有的 $-\infty < x, y < \infty$, 随机变量 X 和 Y 的联合密度为如下形式:

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} \right] \right\}$$

那么称它们为二元正态分布. 现在我们来计算给定 $Y=y$ 的条件下 X 的条件密度. 为了做到这点, 我们先把与 x 无关的因子都收集一起, 并将它们表示为 C_i, 最后这个常数可利用 $\int_{-\infty}^{\infty} f_{X|Y}(x|y)\,\mathrm{d}x = 1$ 得到.

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f(x,y)}{f_Y(y)} = C_1 f(x,y) \\ &= C_2 \exp\left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\frac{x(y-\mu_y)}{\sigma_x\sigma_y} \right] \right\} \\ &= C_3 \exp\left\{ -\frac{1}{2\sigma_x^2(1-\rho^2)} \left[x^2 - 2x\left(\mu_x + \rho\frac{\sigma_x}{\sigma_y}(y-\mu_y)\right) \right] \right\} \\ &= C_4 \exp\left\{ -\frac{1}{2\sigma_x^2(1-\rho^2)} \left[x - \left(\mu_x + \rho\frac{\sigma_x}{\sigma_y}(y-\mu_y)\right) \right]^2 \right\} \end{aligned}$$

注意到上述函数为正态分布的密度函数, 因此可以推得, 给定 $Y=y$ 的条件下, 随机变量 X 服从正态分布, 期望为 $\mu_x + \rho\frac{\sigma_x}{\sigma_y}(y-\mu_y)$, 方差为 $\sigma_x^2(1-\rho^2)$. 而且, 由于 Y, X 的联合密度同 X, Y 的联合密度的形式是一样的, 只需将 μ_x, σ_x 与 μ_x, σ_y 的地位相互对换即可. 由此可知, 在 $X=x$ 给定之下, Y 的条件分布为正态分布, 期望为 $\mu_y + \rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)$, 方差为 $\sigma_y^2(1-\rho^2)$. 从这些结果可以看出, 对于联合正态的随机变量 X, Y 来说, 它们相互独立的充要条件是 X, Y 的相关系数 $\rho = 0$. (这个结果也可从它们的联合密度直接看出, 因为只有当 $\rho = 0$ 时, 它们的联合密度才能分解成两个因子, 其中一个因子只与 x 有关, 另一个因子只与 y 有关.)

X 的边缘密度可以通过如下计算得到:

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y$$
$$= C\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]\right\}\mathrm{d}y$$

其中 $C = \dfrac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}$, 作变量替换 $w = (y-\mu_y)/\sigma_y$ 得

$$\begin{aligned} f_X(x) &= C\sigma_y \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right\} \\ &\quad\times \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(w^2 - 2\rho\frac{x-\mu_x}{\sigma_x}w\right)\right\}\mathrm{d}w \\ &= C\sigma_y \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x-\mu_x}{\sigma_x}\right)^2(1-\rho^2)\right\} \\ &\quad\times \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(w - \rho\frac{x-\mu_x}{\sigma_x}\right)^2\right\}\mathrm{d}w \end{aligned}$$

由

$$\frac{1}{\sqrt{2\pi(1-\rho^2)}}\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(w - \frac{\rho}{\sigma_x}(x-\mu_x)\right)^2\right\}\mathrm{d}w = 1$$

可以得到

$$f_X(x) = C\sigma_y\sqrt{2\pi(1-\rho^2)}\mathrm{e}^{-(x-\mu_x)^2/2\sigma_x^2} = \frac{1}{\sqrt{2\pi}\sigma_x}\mathrm{e}^{-(x-\mu_x)^2/2\sigma_x^2}$$

也即, X 服从正态分布, 均值为 μ_x, 方差为 σ_x^2. 类似地, Y 也服从正态分布, 均值为 μ_y, 方差为 σ_y^2. ■

例 5d 考虑 $n+m$ 次重复试验, 每次成功的概率相同. 然而, 假设此概率并不是固定的, 而是一个随机变量, 其分布为 $(0,1)$ 上的均匀分布. 试问已知 $n+m$ 次试验有 n 次成功的条件下每次成功的概率的条件分布?

解: 如果令 X 表示试验成功概率, 那么 X 为 $(0,1)$ 上均匀随机变量. 同样, 给定 $X=x$ 条件下, $n+m$ 次试验是独立重复试验, 其成功的概率为 x, 而且成功的次数 N 为参数为 $(n+m, x)$ 的二项随机变量. 因此, 给定 $N=n$ 的条件下 X 的条件密度如下:

$$\begin{aligned} f_{X|N}(x|n) &= \frac{P\{N=n|X=x\}f_X(x)}{P\{N=n\}} = \frac{\dbinom{n+m}{n}x^n(1-x)^m}{P\{N=n\}} \qquad 0<x<1 \\ &= cx^n(1-x)^m \end{aligned}$$

其中 c 不依赖于 x. 因此, 其分布密度是 β 分布, 参数为 $(n+1, m+1)$.

上述结果很有意义, 它阐述了若重复试验的每次成功的概率 (先验概率) 为 $(0,1)$ 上均匀分布 [或者, 也就是参数为 $(1,1)$ 的 β 分布], 则经过 $n+m$ 次试验后, 总共有 n 次试验成功的条件下, 其后验分布 (或条件分布) 是参数为 $(1+n, 1+m)$

的 β 分布. 这点很有价值, 它加深了我们的对 β 分布的直观认识. ■

*6.6　次序统计量

设 $X_1, X_2, \cdots, X_n$ 为 n 个独立同分布的连续型随机变量, 其公共分布函数为 $F(x)$, 密度函数为 $f(x)$. 定义

$$\begin{aligned} X_{(1)} &= X_1, X_2, \cdots, X_n \text{ 中的最小者} \\ X_{(2)} &= X_1, X_2, \cdots, X_n \text{ 中的第二小者} \\ &\vdots \\ X_{(j)} &= X_1, X_2, \cdots, X_n \text{ 中的第} j \text{ 小者} \\ &\vdots \\ X_{(n)} &= X_1, X_2, \cdots, X_n \text{ 中的最大者} \end{aligned}$$

排序后的 $X_{(1)} \leqslant X_{(2)} \leqslant \cdots \leqslant X_{(n)}$ 称为 $X_1, X_2, \cdots, X_n$ 的**次序统计量**(order statistics). 换言之, $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ 是将 $X_1, \cdots, X_n$ 排序以后的值.

现在求 $X_{(1)}, \cdots, X_{(n)}$ 的联合密度. $X_{(1)}, \cdots, X_{(n)}$ 的取值为 $x_1 < x_2 < \cdots < x_n$ 的充要条件是存在 $(1, 2, \cdots, n)$ 的一个排列 $(i_1, \cdots, i_n)$, 使得

$$X_1 = x_{i_1}, X_2 = x_{i_2}, \cdots, X_n = x_{i_n}$$

而对于任何 $(1, 2, \cdots, n)$ 的排列 $(i_1, \cdots, i_n)$,

$$\begin{aligned} &P\left\{x_{i_1} - \frac{\varepsilon}{2} < X_1 < x_{i_1} + \frac{\varepsilon}{2}, \cdots, x_{i_n} - \frac{\varepsilon}{2} < X_n < x_{i_n} + \frac{\varepsilon}{2}\right\} \\ \approx\ &\varepsilon^n f_{X_1,\cdots,X_n}(x_{i_1}, \cdots, x_{i_n}) = \varepsilon^n f(x_{i_1}) \cdots f(x_{i_n}) = \varepsilon^n f(x_1) \cdots f(x_n) \end{aligned}$$

由此可知

$$\begin{aligned} &P\left\{x_1 - \frac{\varepsilon}{2} < X_{(1)} < x_1 + \frac{\varepsilon}{2}, \cdots, x_n - \frac{\varepsilon}{2} < X_{(n)} < x_n + \frac{\varepsilon}{2}\right\} \\ =&P\left\{\bigcup_{(i_1,\cdots,i_n)} \left\{x_{i_1} - \frac{\varepsilon}{2} < X_1 < x_{i_1} + \frac{\varepsilon}{2}, \cdots, x_{i_n} - \frac{\varepsilon}{2} < X_n < x_{i_n} + \frac{\varepsilon}{2}\right\}\right\} \\ =&\sum_{(i_1,\cdots,i_n)} P\left\{x_{i_1} - \frac{\varepsilon}{2} < X_1 < x_{i_1} + \frac{\varepsilon}{2}, \cdots, x_{i_n} - \frac{\varepsilon}{2} < X_n < x_{i_n} + \frac{\varepsilon}{2}\right\} \\ \approx&\sum_{(i_1,\cdots,i_n)} \varepsilon^2 f(x_{i_1}) \cdots f(x_{i_n}) = n!\varepsilon^n f(x_1) \cdots f(x_n) \end{aligned}$$

上式两端除以 ε^n 并令 $\varepsilon \to 0$, 得

$$f_{X_{(1)},\cdots,X_{(n)}}(x_1, x_2, \cdots, x_n) = n!\, f(x_1) \cdots f(x_n) \qquad x_1 < x_2 < \cdots < x_n \tag{6.1}$$

我们可以这样直观地解释 (6.1), 向量 $\langle X_{(1)}, \cdots, X_{(n)}\rangle$ 等于 $\langle x_1, \cdots, x_n\rangle$ 的充要条件是 $\langle X_1, \cdots, X_n\rangle$ 等于 $\langle x_1, \cdots, x_n\rangle$ 的 $n!$ 种排列之一. 而 $\langle X_1, \cdots, X_n\rangle$ 等于 $\langle x_1, \cdots, x_n\rangle$ 的任一排列的概率 (密度) 刚好是 $f(x_1)\cdots f(x_n)$. 由此, 可知 (6.1) 式成立.

例 6a 有三个人"随机地分布"在长为一英里的一段路上. 找出没有两个人之间距离小于 d 英里的概率 ($d \leqslant 1/2$).

解: 我们假定"随机地分布"是指三个人相互独立地均匀分布在一英里的路段上. 用 X_i 表示第 i 个人在路上的位置. 所求的概率是 $P\{X_{(i)} > X_{(i-1)} + d, i = 2,3\}$. $X_{(1)}, X_{(2)}, X_{(3)}$ 的联合密度为

$$f_{X_{(1)},X_2,X_3}(x_1,x_2,x_3) = 3! \qquad 0 < x_1 < x_2 < x_3 < 1$$

由密度与概率的关系可导出

$$P\{X_{(i)} > X_{(i-1)} + d, i = 2,3\} = \iiint\limits_{\substack{x_i > x_{i-1}+d \\ i=2,3}} f_{X_{(1)},X_{(2)},X_{(3)}}(x_1,x_2,x_3)\,\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}x_3$$

$$=3!\int_0^{1-2d}\int_{x_1+d}^{1-d}\int_{x_2+d}^{1}\mathrm{d}x_3\,\mathrm{d}x_2\,\mathrm{d}x_1 = 6\int_0^{1-2d}\int_{x_1+d}^{1-d}(1-d-x_2)\mathrm{d}x_2\,\mathrm{d}x_1$$

$$=6\int_0^{1-2d}\int_0^{1-2d-x_1} y_2\,\mathrm{d}y_2\,\mathrm{d}x_1$$

其中, 我们作了变量替换 $y_2 = 1 - d - x_2$. 将等式继续下去, 得到

$$= 3\int_0^{1-2d}(1-2d-x_1)^2\mathrm{d}x_1 = 3\int_0^{1-2d} y_1^2\,\mathrm{d}y_1 = (1-2d)^3$$

这样, 当 $d \leqslant 1/2$ 时, 一英里的路段上随机地分布的三个人之间两两最小距离大于 d 的概率为 $(1-2d)^3$. 利用这个方法还可以证明一英里的路段上随机地分布的 n 个人之间两两最小距离大于 d 的概率为

$$[1-(n-1)d]^n \qquad d \leqslant \frac{1}{n-1}$$

其证明作为练习. ■

将 (6.1) 式进行积分可求得第 j 个次序统计量 $X_{(j)}$ 的密度函数. 但也可以用下列方法直接推得. $X_{(j)} = x$ 意味着 $X_1, \cdots, X_n$ 中有 $j-1$ 个值小于 x, 有一个值等于 x, 有 $n-j$ 个值大于 x. 对于给定的一个变量等于 x, 给定的 $j-1$ 个变量的值小于 x, 其余的变量的值大于 x 的概率密度为

$$[F(x)]^{j-1}[1-F(x)]^{n-j}f(x)$$

由于 n 个变量分成三个组的方法有 $\binom{n}{j-1, n-j, 1} = \dfrac{n!}{(n-j)!(j-1)!}$ 种. 这样,

$X_{(j)}$ 的密度函数为

$$f_{X_{(j)}}(x) = \frac{n!}{(n-j)!(j-1)!}[F(x)]^{j-1}[1-F(x)]^{n-j}f(x) \tag{6.2}$$

例 6b　设 $X_1, \cdots, X_{2n+1}$ 为独立同分布的随机变量 (统计上称 $X_1, \cdots, X_{2n+1}$ 为一个大小为 $2n+1$ 的样本). 次序统计量 $X_{(n+1)}$ 称为样本中位数. 现设 X_1, X_2, X_3 为 $(0,1)$ 上均匀分布的一组样本. 求样本中位数落入区间 $(1/4, 3/4)$ 的概率.

解: 由样本中位数的定义可知, 本例的样本中位数就是 $X_{(2)}$. 利用 (6.2) 式, $X_{(2)}$ 的密度函数

$$f_{X_{(2)}}(x) = \frac{3!}{1!1!}x(1-x) \qquad 0 < x < 1$$

因此所求概率为

$$P\left\{\frac{1}{4} < X_{(2)} < \frac{3}{4}\right\} = 6\int_{1/4}^{3/4} x(1-x)\,\mathrm{d}x = 6\left\{\frac{x^2}{2} - \frac{x^3}{3}\right\}\Big|_{x=1/4}^{x=3/4} = \frac{11}{16} \qquad \blacksquare$$

只需将 (6.2) 式积分就可得到 $X_{(j)}$ 的分布函数

$$F_{X_{(j)}}(y) = \frac{n!}{(n-j)!(j-1)!}\int_{-\infty}^{y}[F(x)]^{j-1}[1-F(x)]^{n-j}f(x)\,\mathrm{d}x \tag{6.3}$$

但是也可用其他方法求得 $F_{X_{(j)}}(y)$, 注意到第 j 个次序统计量小于或等于 y 的充要条件是 $X_1, \cdots, X_n$ 中小于或等于 y 的个数等于 j 或比 j 更多, 即

$$\begin{aligned} F_{X_{(j)}}(y) = P\{X_{(j)} \leqslant y\} &= P\{j\text{个或更多}X_i \leqslant y\} \\ &= \sum_{k=j}^{n}\binom{n}{k}[F(y)]^k[1-F(y)]^{n-k} \end{aligned} \tag{6.4}$$

利用分部积分法, 可以证明 (6.3) 式和 (6.4) 式右边的两式是恒等的. 如果令 F 为 $(0,1)$ 上的均匀分布 [即 $f(x) = 1, 0 < x < 1$], 比较 (6.3) 式和 (6.4) 式可得到一个十分有用的分析恒等式

$$\sum_{k=j}^{n}\binom{n}{k}y^k(1-y)^{n-k} = \frac{n!}{(n-j)!(j-1)!}\int_0^y x^{j-1}(1-x)^{n-j}\mathrm{d}x \qquad 0 \leqslant y \leqslant 1 \tag{6.5}$$

类似于公式 (6.2) 的推导, 我们可以求得 $X_{(i)}$ 与 $X_{(j)}$ 的联合分布密度 $(i < j)$:

$$\begin{aligned} f_{X_{(i)},X_{(j)}}(x_i, x_j) =& \frac{n!}{(i-1)!(j-i-1)!(n-j)!}[F(x_i)]^{i-1} \\ &\times [F(x_j) - F(x_i)]^{j-i-1}[1-F(x_j)]^{n-j}f(x_i)f(x_j) \quad x_i < x_j \end{aligned} \tag{6.6}$$

例 6c (随机样本极差的分布)　设 $X_1, \cdots, X_n$ 为 n 个独立同分布的随机变量. 随机变量 $R = X_{(n)} - X_{(1)}$ 称为极差. 设 X_i 的分布函数为 $F(x)$, 相应的密度函数

为 $f(x)$, 则 R 的分布可通过 (6.6) 式求得. 对于 $a \geqslant 0$,

$$P\{R \leqslant a\} = P\{X_{(n)} - X_{(1)} \leqslant a\} = \iint\limits_{x_n - x_1 \leqslant a} f_{X_{(1)},X_{(n)}}(x_1, x_n)\,\mathrm{d}x_1\,\mathrm{d}x_n$$

$$= \int_{-\infty}^{\infty}\int_{x_1}^{x_1+a} \frac{n!}{(n-2)!}[F(x_n) - F(x_1)]^{n-2} f(x_1) f(x_n)\,\mathrm{d}x_n\,\mathrm{d}x_1$$

作变量替换 $y = F(x_n) - F(x_1), \mathrm{d}y = f(x_n)\,\mathrm{d}x_n$, 得

$$\int_{x_1}^{x_1+a} [F(x_n) - F(x_1)]^{n-2} f(x_n)\,\mathrm{d}x_n = \int_0^{F(x_1+a)-F(x_1)} y^{n-2}\,\mathrm{d}y$$

$$= \frac{1}{n-1}[F(x_1+a) - F(x_1)]^{n-1}$$

因此,

$$P\{R \leqslant a\} = n\int_{-\infty}^{\infty} [F(x_1+a) - F(x_1)]^{n-1} f(x_1)\,\mathrm{d}x_1 \tag{6.7}$$

(6.7) 式右端的积分只在很少几种情况下才能积出来, 写成显示表达式. 当 X_i 的分布为 $(0,1)$ 上均匀分布时, R 的分布可由 (6.7) 式算出来. 对于 $a \in (0,1)$,

$$P\{R < a\} = n\int_0^1 [F(x_1+a) - F(x_1)]^{n-1} f(x_1)\,\mathrm{d}x_1$$

$$= n\int_0^{1-a} a^{n-1}\mathrm{d}x_1 + n\int_{1-a}^{1} (1-x_1)^{n-1}\mathrm{d}x_1 = n(1-a)a^{n-1} + a^n$$

将上式求微商可求得极差的密度函数

$$f_R(a) = \begin{cases} n(n-1)a^{n-2}(1-a) & 0 \leqslant a \leqslant 1 \\ 0 & \text{其他} \end{cases}$$

即独立的 $(0,1)$ 上均匀随机变量序列的极差的分布是 β 分布, 参数为 $(n-1, 2)$. ■

6.7 随机变量函数的联合分布

设 X_1, X_2 具有联合密度函数 f_{X_1,X_2}. 有时我们需要求出 Y_1, Y_2 的联合分布, 其中 Y_1, Y_2 为 X_1, X_2 的函数. 具体地说, 设 $Y_1 = g_1(X_1, X_2), Y_2 = g_2(X_1, X_2)$, 并且函数 g_1, g_2 满足下列条件:

1. 由下列方程组

$$y_1 = g_1(x_1, x_2)$$

$$y_2 = g_2(x_1, x_2)$$

可唯一地解出 x_1, x_2 来, 即求出 $x_1 = h_1(y_1, y_2), x_2 = h_2(y_1, y_2)$.

2. 函数 g_1, g_2 对一切 (x_1, x_2) 具有连续偏导数, 并且下面的 2×2 行列式对一

切 (x_1, x_2) 不为 0

$$J(x_1, x_2) = \begin{vmatrix} \dfrac{\partial g_1}{\partial x_1} & \dfrac{\partial g_1}{\partial x_2} \\ \dfrac{\partial g_2}{\partial x_1} & \dfrac{\partial g_2}{\partial x_2} \end{vmatrix} \equiv \frac{\partial g_1}{\partial x_1}\frac{\partial g_2}{\partial x_2} - \frac{\partial g_1}{\partial x_2}\frac{\partial g_2}{\partial x_1} \neq 0$$

在这两条件之下, 可以证明 Y_1, Y_2 的联合密度函数可通过下式得到:

$$f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}(x_1, x_2)\left|J(x_1, x_2)\right|^{-1} \tag{7.1}$$

其中 $x_1 = h_1(y_1, y_2), x_2 = h_2(y_1, y_2)$.

(7.1) 式的证明可从下式入手

$$P\{Y_1 \leqslant y_1, Y_2 \leqslant y_2\} = \iint\limits_{\substack{(x_1,x_2):\\ g_1(x_1,x_2)\leqslant y_1\\ g_2(x_1,x_2)\leqslant y_2}} f_{X_1,X_2}(x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 \tag{7.2}$$

Y_1, Y_2 的联合密度函数可通过将上式对 y_1, y_2 求偏微商得到. 微商的结果刚好等于 (7.1) 式的右边. 微商的过程作为高等微积分的一个练习本书不再细述. 下面通过例子来熟悉相应的计算过程.

例 7a 设 X_1, X_2 为联合连续的随机变量, 其联合密度函数为 f_{X_1,X_2}. 令 $Y_1 = X_1 + X_2, Y_2 = X_1 - X_2$. 求出 Y_1, Y_2 的联合密度函数.

解: 令 $g_1(x_1, x_2) = x_1 + x_2, g_2(x_1, x_2) = x_1 - x_2$, 经计算

$$J(x_1, x_2) = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = -2$$

由 $y_1 = x_1 + x_2, y_2 = x_1 - x_2$ 解得 $x_1 = (y_1 + y_2)/2, x_2 = (y_1 - y_2)/2$. 利用 (7.1) 式可得

$$f_{Y_1,Y_2}(y_1, y_2) = \frac{1}{2} f_{X_1,X_2}\big((y_1 + y_2)/2, (y_1 - y_2)/2\big)$$

例如, X_1, X_2 为独立同分布的 $(0,1)$ 均匀随机变量, 则

$$f_{Y_1,Y_2}(y_1, y_2) = \begin{cases} \dfrac{1}{2} & 0 \leqslant y_1 + y_2 \leqslant 2, 0 \leqslant y_1 - y_2 \leqslant 2 \\ 0 & \text{其他} \end{cases}$$

若 X_1, X_2 为相互独立的指数随机变量, 其相应的参数为 λ_1, λ_2, 则

$$f_{Y_1,Y_2}(y_1, y_2) = \begin{cases} \dfrac{\lambda_1\lambda_2}{2} \exp\left\{-\lambda_1\left(\dfrac{y_1+y_2}{2}\right) - \lambda_2\left(\dfrac{y_1-y_2}{2}\right)\right\} & y_1+y_2 \geqslant 0, y_1-y_2 \geqslant 0 \\ 0 & \text{其他} \end{cases}$$

最后, 若 X_1, X_2 为相互独立的标准正态随机变量, 则

$$\begin{aligned} f_{Y_1,Y_2}(y_1,y_2) &= \frac{1}{4\pi}\mathrm{e}^{-[(y_1+y_2)^2/8+(y_1-y_2)^2/8]} \\ &= \frac{1}{4\pi}\mathrm{e}^{-(y_1^2+y_2^2)/4} = \frac{1}{\sqrt{4\pi}}\mathrm{e}^{-y_1^2/4}\frac{1}{\sqrt{4\pi}}\mathrm{e}^{-y_2^2/4} \end{aligned}$$

我们不仅得知 X_1+X_2 与 X_1-X_2 为正态随机变量, 期望为 0, 方差为 2(这一点与命题 3.2 的结论相同), 而且, X_1+X_2 与 X_1-X_2 还是相互独立的. (事实上, 我们还可以指出若 X_1, X_2 独立同分布, 其分布函数为 F, 则 X_1+X_2 与 X_1-X_2 相互独立的充要条件为 F 是正态分布函数.)

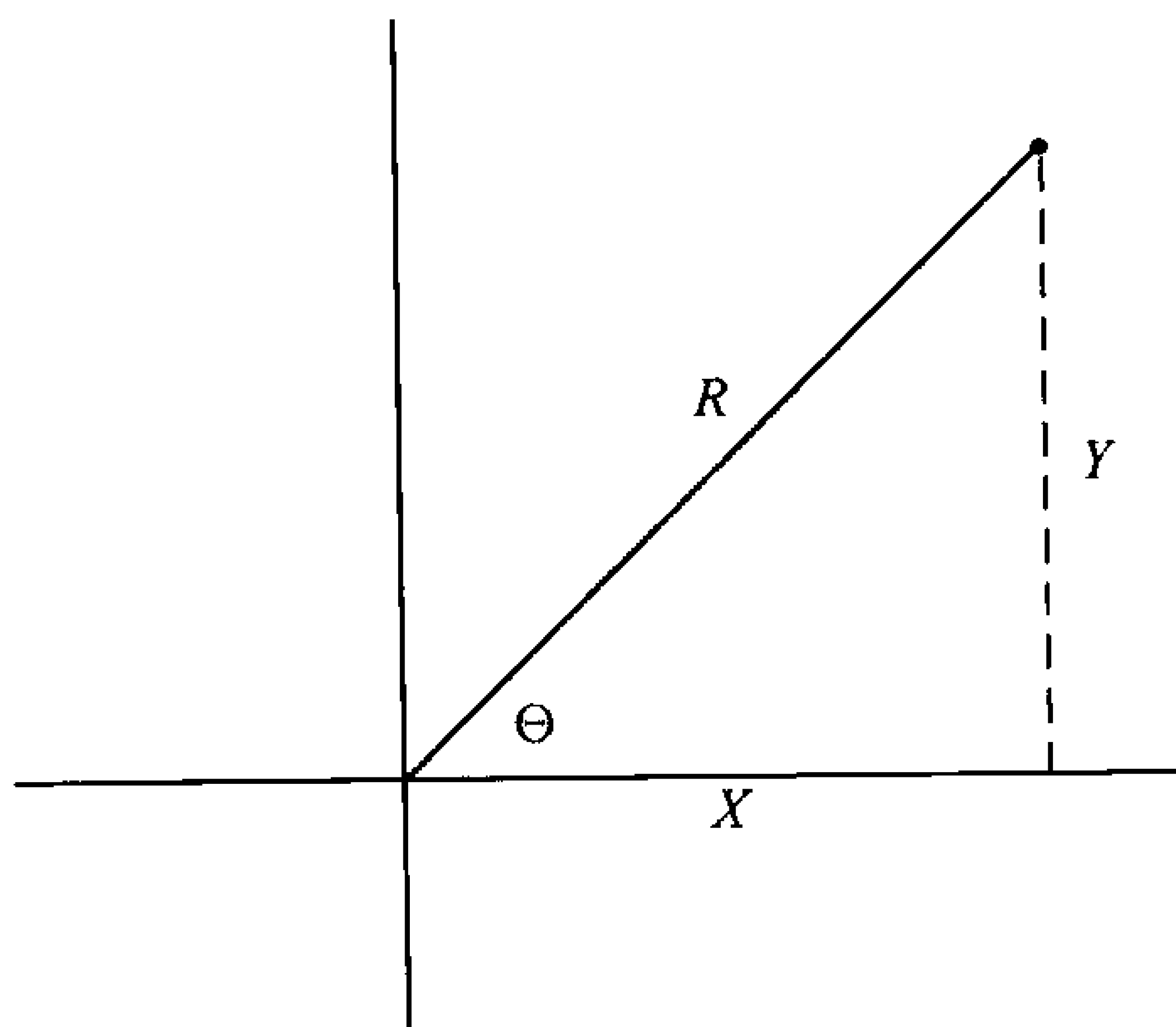


图 6.4 随机点 $(X,Y)=(R,\Theta)$

例 7b 令 (X,Y) 表示平面上一个随机点, 并假设其直角坐标 X 和 Y 是相互独立的标准正态随机变量. 现在感兴趣的是这个随机点的极坐标表示 R,Θ 的联合分布 (见图 6.4).

令 $r=g_1(x,y)=\sqrt{x^2+y^2}, \theta=g_2(x,y)=\tan^{-1}(y/x)$,

$$\frac{\partial g_1}{\partial x}=\frac{x}{\sqrt{x^2+y^2}} \qquad \frac{\partial g_1}{\partial y}=\frac{y}{\sqrt{x^2+y^2}}$$

$$\frac{\partial g_2}{\partial x}=\frac{1}{1+(y/x)^2}\left(\frac{-y}{x^2}\right)=\frac{-y}{x^2+y^2} \qquad \frac{\partial g_2}{\partial y}=\frac{1}{x[1+(y/x)^2]}=\frac{x}{x^2+y^2}$$

由此得

$$J(x,y)=\frac{x^2}{(x^2+y^2)^{3/2}}+\frac{y^2}{(x^2+y^2)^{3/2}}=\frac{1}{\sqrt{x^2+y^2}}=\frac{1}{r}$$

由于 X,Y 的联合密度函数为

$$f(x,y)=\frac{1}{2\pi}\mathrm{e}^{-(x^2+y^2)/2}$$

现在求变换 $r=\sqrt{x^2+y^2}, \theta=\tan^{-1}(y/x)$ 之下 R,Θ 的联合密度, 为了求得 (R,Θ) 的联合密度, 我们将 (x,y) 平面分成 4 大块, $(x>0,y>0),(x<0,y>0),(x<0,y<0)$ 和 $(x>0,y<0)$, 而 $r=\sqrt{x^2+y^2}, \theta=\tan^{-1}(y/x)$ 将 $(x>0,y>0)$ 变换成 $0<\theta<\dfrac{\pi}{2}, 0<r<\infty$. 现在考虑条件密度

$$f(x,y|X>0,Y>0)=\frac{f(x,y)}{P(X>0,Y>0)}=\frac{2}{\pi}\mathrm{e}^{-(x^2+y^2)/2}, x>0,y>0$$

在变换 $R=\sqrt{X^2+Y^2}$ 和 $\Theta=\tan^{-1}(Y|X)$ 之下, (R,Θ) 的条件密度为

$$f(r,\theta|X>0,Y>0)=\frac{2}{\pi}r\mathrm{e}^{-\frac{r^2}{2}}, \quad 0<\theta<\pi/2, \quad 0<r<\infty.$$

类似地, 我们可以得到

$$f(r,\theta|X<0,Y>0)=\frac{2}{\pi}re^{-\frac{r^2}{2}},\quad \pi/2<\theta<\pi,\quad 0<r<\infty,$$
$$f(r,\theta|X<0,Y>0)=\frac{2}{\pi}re^{-\frac{r^2}{2}},\quad \pi<\theta<3\pi/2,\quad 0<r<\infty$$
$$f(r,\theta|X<0,Y<0)=\frac{2}{\pi}re^{-r^2/2},\quad 3\pi/2<\theta<2\pi,\quad 0<r<\infty$$

而 (R,Θ) 的联合密度刚好是这四个条件密度的等权加权平均 $\left(\text{权为 }\frac{1}{4}\right)$, 即 $R=\sqrt{X^2+Y^2},\Theta=\arctan(Y/X)$ 的联合密度函数为

$$f(r,\theta)=\frac{1}{2\pi}re^{-r^2/2}\qquad 0<\theta<2\pi,0<r<\infty$$

这个联合密度函数可分解为两个密度函数的乘积, 因此, R 与 Θ 是相互独立的. Θ 是 $(0,2\pi)$ 上的均匀随机变量, 而 R 的分布是著名的瑞利分布, 其密度函数为

$$f(r)=re^{-r^2/2},\quad 0<r<\infty$$

(例如, 当我们关心平面打靶问题时, 如果水平和垂直误差是相互独立的标准正态分布, 则弹着点相对于靶心的方位角不仅仅是均匀分布, 而且与弹着点离靶心的距离相互独立.)

上面得到的结果是十分有意义的, 因为极坐标的分布与直角坐标的这种关系不是一眼就能看出来的.

如果我们希望求出 R^2 和 Θ 的联合分布, 变换

$$d=g_1(x,y)=x^2+y^2\qquad \theta=g_2(x,y)=\arctan(y/x)$$

的雅可比行列式为

$$J=\begin{vmatrix} 2x & 2y \\ \dfrac{-y}{x^2+y^2} & \dfrac{x}{x^2+y^2}\end{vmatrix}=2$$

利用公式 (7.1) 可得

$$f(d,\theta)=\frac{1}{2}e^{-d/2}\frac{1}{2\pi}\qquad 0<d<\infty,0<\theta<2\pi$$

由此看出 R^2 和 Θ 相互独立, R^2 具有指数分布, 参数为 1/2. 但 $R^2=X^2+Y^2$, 由定义可知, R^2 具有 χ^2 分布, 自由度为 2. 这样, 我们验证了自由度为 2 的 χ^2 分布与参数为 1/2 的指数分布是相同的.

上面的结果可用于模拟标准正态随机变量. 令 U_1,U_2 为相互独立的均匀随机数 (服从 (0, 1) 上的均匀分布), 我们将把 U_1,U_2 转化为两独立的标准正态随机变量 X_1 和 X_2. 在求正态随机变量之前, 先求出 (X_1,X_2) 的极坐标表示 (R,Θ). 利

用 (X,Y) 与 (R,Θ) 之间的关系可知 R^2 与 Θ 相互独立. $R^2 = X^2+Y^2$ 具有指数分布, 参数为 $\lambda = 1/2$. 但 $-2\ln U_1$ 具有这样的分布, 对于 $x>0$, 有

$$P\{-2\ln U_1 < x\} = P\{\ln U_1 > -x/2\} = P\{U_1 > \mathrm{e}^{-x/2}\} = 1-\mathrm{e}^{-x/2}$$

另一方面, $2\pi U_2$ 为 $(0,2\pi)$ 上的均匀随机变量, 利用它可生成 Θ. 如果令

$$R^2 = -2\ln U_1 \qquad \Theta = 2\pi U_2$$

R^2 可以看成 (X_1,X_2) 到原点的距离的平方, Θ 就是 (X_1,X_2) 的方位角. 由 $X_1 = R\cos\Theta, X_2 = R\sin\Theta$, 我们得到

$$X_1 = \sqrt{-2\ln U_1}\cos(2\pi U_2) \qquad X_2 = \sqrt{-2\ln U_1}\sin(2\pi U_2)$$

是相互独立的标准正态随机变量. ■

例 7c 设 X 和 Y 为相互独立的 Γ 随机变量, 其参数分别为 (α,λ) 和 (β,λ), 计算 $U=X+Y$ 和 $V=X/(X+Y)$ 的联合分布.

解: X 和 Y 的联合密度为

$$\begin{aligned} f_{X,Y}(x,y) &= \frac{\lambda\mathrm{e}^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)}\frac{\lambda\mathrm{e}^{-\lambda y}(\lambda y)^{\beta-1}}{\Gamma(\beta)} \\ &= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)}\mathrm{e}^{-\lambda(x+y)}x^{\alpha-1}y^{\beta-1} \qquad x>0, y>0 \end{aligned}$$

令 $g_1(x,y) = x+y, g_2(x,y) = x/(x+y)$, 可得

$$\frac{\partial g_1}{\partial x} = \frac{\partial g_1}{\partial y} = 1 \qquad \frac{\partial g_2}{\partial x} = \frac{y}{(x+y)^2} \qquad \frac{\partial g_2}{\partial y} = -\frac{x}{(x+y)^2}$$

因此

$$J(x,y) = \begin{vmatrix} 1 & 1 \\ \dfrac{y}{(x+y)^2} & \dfrac{-x}{(x+y)^2} \end{vmatrix} = -\frac{1}{x+y}$$

由方程组 $u=x+y, v=x/(x+y)$, 得到解 $x=uv, y=u(1-v)$. 利用公式 (7.1),

$$f_{U,V}(u,v) = f_{X,Y}[uv, u(1-v)]u = \frac{\lambda\mathrm{e}^{-\lambda u}(\lambda u)^{\alpha+\beta-1}}{\Gamma(\alpha+\beta)}\frac{v^{\alpha-1}(1-v)^{\beta-1}\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$$

由上式可知 $X+Y$ 与 $X/(X+Y)$ 相互独立. $X+Y$ 具有 Γ 分布, 参数为 $(\alpha+\beta,\lambda)$. $X/(X+Y)$ 具有 β 分布, 参数为 (α,β). 由上式还可以看出 β 分布中的归一化因子 $B(\alpha,\beta)$ 为

$$B(\alpha,\beta) = \int_0^1 v^{\alpha-1}(1-v)^{\beta-1}\,\mathrm{d}v = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

上面得到的结果很有意思. 比如, 假设共有 $n+m$ 项工作需要完成, 完成每项工作

所需的时间服从指数分布, 并且完成各项工作所需时间是相互独立的. 现在假定将这 $n+m$ 项工作分配给两人去完成. 甲完成其中 n 件, 乙完成余下的 m 件. 甲乙两人所花费时间分别为 X,Y. X,Y 相互独立, 并且其分布为 Γ 分布, 参数分别为 (n,λ) 和 (m,λ). 甲所花费的时间占总任务时间 (即 $X+Y$) 的比例具有 β 分布, 参数为 (n,m). ■

现在设 $X_1,\cdots,X_n$ 的联合密度已经给出, 我们希望求得 $Y_1,\cdots,Y_n$ 的联合密度函数, 其中

$$Y_1=g_1(X_1,\cdots,X_n)\quad Y_2=g_2(X_1,\cdots,X_n)\quad\cdots\quad Y_n=g_n(X_1,\cdots,X_n)$$

其解法与二维随机变量的函数的密度的求法类似. 我们假定 g_i 具有连续偏导数并且雅可比行列式 $J(x_1,\cdots,x_n)\neq 0$ 对一切 $(x_1,\cdots,x_n)$ 成立, 其中

$$J(x_1,\cdots,x_n)=\begin{vmatrix}\dfrac{\partial g_1}{\partial x_1} & \dfrac{\partial g_1}{\partial x_2} & \cdots & \dfrac{\partial g_1}{\partial x_n}\\ \dfrac{\partial g_2}{\partial x_1} & \dfrac{\partial g_2}{\partial x_2} & \cdots & \dfrac{\partial g_2}{\partial x_n}\\ \cdots & \cdots & \cdots & \cdots\\ \dfrac{\partial g_n}{\partial x_1} & \dfrac{\partial g_n}{\partial x_2} & \cdots & \dfrac{\partial g_n}{\partial x_n}\end{vmatrix}$$

进一步, 我们假定方程组

$$\begin{aligned}y_1&=g_1(x_1,\cdots,x_n)\\ y_2&=g_2(x_1,\cdots,x_n)\\ &\cdots\\ y_n&=g_n(x_1,\cdots,x_n)\end{aligned}$$

具有唯一解, 即 $x_1=h_1(y_1,\cdots,y_n),\cdots,x_n=h_n(y_1,\cdots,y_n)$. 在这些假定之下, $Y_1,\cdots,Y_n$ 的联合密度为

$$f_{Y_1,\cdots,Y_n}(y_1,\cdots,y_n)=f_{X_1,\cdots,X_n}(x_1,\cdots,x_n)\left|J(x_1,\cdots,x_n)\right|^{-1}\tag{7.3}$$

其中 $x_i=h_i(y_1,\cdots,y_n), i=1,2,\cdots,n$.

例 7d 设 X_1,X_2 和 X_3 为相互独立的标准正态随机变量, 令 $Y_1=X_1+X_2+X_3$, $Y_2=X_1-X_2$, $Y_3=X_1-X_3$. 计算 Y_1,Y_2,Y_3 的联合密度.

解: 计算 Y_1,Y_2,Y_3 相对于 X_1,X_2,X_3 的雅可比行列式,

$$J=\begin{vmatrix}1 & 1 & 1\\ 1 & -1 & 0\\ 1 & 0 & -1\end{vmatrix}=3$$

由 Y_i 的表达式可解得

$$X_1 = \frac{1}{3}(Y_1 + Y_2 + Y_3) \quad X_2 = \frac{1}{3}(Y_1 - 2Y_2 + Y_3) \quad X_3 = \frac{1}{3}(Y_1 + Y_2 - 2Y_3)$$

利用 (7.3) 式可得

$$f_{Y_1,Y_2,Y_3}(y_1,y_2,y_3) = \frac{1}{3} f_{X_1,X_2,X_3}\big((y_1+y_2+y_3)/3, (y_1-2y_2+y_3)/3, (y_1+y_2-2y_3)/3\big)$$

由于

$$f_{X_1,X_2,X_3}(x_1,x_2,x_3) = \frac{1}{(\sqrt{2\pi})^3}\exp\Big\{-\sum_{i=1}^{3} x_i^2/2\Big\}$$

可得

$$f_{Y_1,Y_2,Y_3}(y_1,y_2,y_3) = \frac{1}{3(\sqrt{2\pi})^3}\exp\{-Q(y_1,y_2,y_3)/2\}$$

其中

$$\begin{aligned} Q(y_1,y_2,y_3) &= \left(\frac{y_1+y_2+y_3}{3}\right)^2 + \left(\frac{y_1-2y_2+y_3}{3}\right)^2 + \left(\frac{y_1+y_2-2y_3}{3}\right)^2 \\ &= \frac{1}{3}y_1^2 + \frac{2}{3}y_2^2 + \frac{2}{3}y_3^2 - \frac{2}{3}y_2y_3 \end{aligned}$$

■

例 7e 令 $X_1, X_2, \cdots, X_n$ 为独立同分布的指数随机变量, 其参数为 λ, 令

$$Y_i = X_1 + X_2 + \cdots + X_i \qquad i = 1, \cdots, n$$

(a) 求 $Y_1, \cdots, Y_n$ 的联合密度函数. (b) 利用 (a) 的结果, 求 Y_n 的密度函数.

解: (a) 变换 $Y_1 = X_1, Y_2 = X_1 + X_2, \cdots, Y_n = X_1 + \cdots + X_n$ 的雅可比行列式为

$$J = \begin{vmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 1 & 1 & 1 & \cdots & 1 \end{vmatrix}$$

若将该行列式展开, 只有第一项非 0, 因此 $J = 1$. 而 $X_1, \cdots, X_n$ 的联合密度为

$$f_{X_1,\cdots,X_n}(x_1,\cdots,x_n) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda(x_1+\cdots+x_n)} \qquad 0 < x_i < \infty, i = 1, \cdots, n$$

将方程组 $Y = X_1, \cdots, Y_n = X_1 + \cdots + X_n$ 反解得

$$X_1 = Y_1, X_2 = Y_2 - Y_1, \cdots, X_i = Y_i - Y_{i-1}, \cdots, X_n = Y_n - Y_{n-1}$$

最后利用 (7.3) 式, 得到 $Y_1, \cdots, Y_n$ 的联合密度函数为

$$\begin{aligned}
&f_{Y_1,\cdots,Y_n}(y_1,y_2,\cdots,y_n)\\
=&f_{X_1,\cdots,X_n}(y_1,y_2-y_1,\cdots,y_i-y_{i-1},\cdots,y_n-y_{n-1})\\
=&\lambda^n\exp\left\{-\lambda\left[y_1+\sum_{i=2}^{n}(y_i-y_{i-1})\right]\right\}\\
=&\lambda^n \mathrm{e}^{-\lambda y_n}\qquad 0<y_1,0<y_i-y_{i-1},i=2,\cdots,n\\
=&\lambda^n \mathrm{e}^{-\lambda y_n}\qquad 0<y_1<y_2<\cdots<y_n
\end{aligned}$$

(b) 为求 Y_n 的边缘密度, 我们必须对联合密度中的其他变量求积分. 现在一个变量一个变量地求积分:

$$f_{Y_2,\cdots,Y_n}(y_2,\cdots,y_n)=\int_0^{y_2}\lambda^n\mathrm{e}^{-\lambda y_n}\,\mathrm{d}y_1=\lambda^n y_2\mathrm{e}^{-\lambda y_n}\qquad 0<y_2<y_3<\cdots<y_n$$

进一步对 y_2 求积分, 得

$$\begin{aligned}
f_{Y_3,\cdots,Y_n}(y_3,\cdots,y_n)&=\int_0^{y_3}\lambda^n y_2\mathrm{e}^{-\lambda y_n}\,\mathrm{d}y_2\\
&=\lambda^n\frac{y_3^2}{2}\mathrm{e}^{-\lambda y_n}\qquad 0<y_3<y_4<\cdots<y_n
\end{aligned}$$

下一步得到

$$f_{Y_4,\cdots,Y_n}(y_4,\cdots,y_n)=\lambda^n\frac{y_4^3}{3\,!}\mathrm{e}^{-\lambda y_n}\qquad 0<y_4<\cdots<y_n$$

继续下去, 最后得到

$$f_{Y_n}(y_n)=\lambda^n\frac{y_n^{n-1}}{(n-1)\,!}\mathrm{e}^{-\lambda y_n}\qquad 0<y_n$$

这个结果与例 3b 的结果是一致的, 即 $X_1+\cdots+X_n$ 是 Γ 随机变量, 参数为 (n,λ).

■

*6.8 可交换随机变量

随机变量序列 $X_1,\cdots,X_n$ 称为可交换的, 如果对于 $1,2,\cdots,n$ 的每一个排列 $i_1,\cdots,i_n$,

$$P\{X_{i_1}\leqslant x_1,X_{i_2}\leqslant x_2,\cdots,X_{i_n}\leqslant x_n\}=P\{X_1\leqslant x_1,X_2\leqslant x_2,\cdots,X_n\leqslant x_n\}$$

对一切 $x_1,\cdots,x_n$ 成立. 换言之, n 个随机变量称为可交换的, 如果它们的联合分布与这些随机变量的次序无关.

当 $X_1,\cdots,X_n$ 为离散型随机变量时, 可交换条件可写成

$$P\{X_{i_1}=x_1,X_{i_2}=x_2,\cdots,X_{i_n}=x_n\}=P\{X_1=x_1,X_2=x_2,\cdots,X_n=x_n\}$$

对一切排列和一切 $x_1,\cdots,x_n$ 成立. 它与下列陈述是等价的: 分布列 $p(x_1,x_2,\cdots,$

$x_n) = P\{X_{i_1} = x_1, X_{i_2} = x_2, \cdots, X_{i_n} = x_n\}$ 是变量 $(x_1, \cdots, x_n)$ 的对称函数, 或者说当变量 $x_1, \cdots, x_n$ 的次序变换以后, 相应的概率值不变.

例 8a 设坛子里一共有 n 个球, 其中 k 个球被认为是特殊的球. 现在从坛子中一个一个地无放回地随机地抽取 n 个球. 记

$$X_i = \begin{cases} 1 & \text{如果第 } i \text{ 次抽到一个特殊的球} \\ 0 & \text{如果抽到的是其他的球} \end{cases}$$

我们将指出 $X_1, X_2, \cdots, X_n$ 是可交换的. 我们需要证明 $p(x_1, \cdots, x_n)$ 是 $(x_1, \cdots, x_n)$ 的对称函数, 对于 $(x_1, \cdots, x_n)$, 只要 $\sum x_i \neq k$, 就必定有 $p(x_1, \cdots, x_n) = 0$. 因此, 我们只需对 $(x_1, \cdots, x_n), \sum_{i=1}^n x_i = k, x_i$等于 0 或 1 证明 $p(x_1, \cdots, x_n)$ 是对称函数即可. 为了对 $p(x_1, \cdots, x_n)$ 有个直观的了解, 取一个具体的向量 $(1, 1, 0, 1, 0, \cdots, 0, 1)$, 其中有 k 个 1, $n-k$ 个 0.

$$p(1,1,0,1,0,\cdots,0,1) = \frac{k}{n}\frac{k-1}{n-1}\frac{n-k}{n-2}\frac{k-2}{n-3}\frac{n-k-1}{n-4}\cdots\frac{1}{2}\frac{1}{1}$$

这个公式比较直观, 由于抽取第一个球是一个特殊的球, 因此, 第一个球是特殊球的概率为 k/n, 在第一个球为特殊球的条件下, 第二个也是特殊球的概率为 $(k-1)/(n-1)$, 在第一、第二均为特殊球的条件下, 第三个为普通球的概率为 $(n-k)/(n-2)$, 如此继续下去, 把这些条件概率乘起来就得到 $p(1,1,0,1,0,\cdots,0,1)$. 这些连乘的分数的分母的连乘积为 $n(n-1)\cdots 1$, 其原因是每抽到一个球, 坛子里的球就少一个. 而分子可表示为两部分的乘积, 一部分为 $k \cdot (k-1) \cdots 1$, 另一部分为 $(n-k) \cdot (n-k-1) \cdots 1$, 这样我们得到

$$p(x_1, \cdots, x_n) = \frac{k\,!\,(n-k)\,!}{n\,!} \qquad x_i = 0, 1, \sum_{i=1}^n x_i = k$$

显然, $p(x_1, \cdots, x_n)$ 是 $(x_1, \cdots, x_n)$ 的对称函数, 即 $(X_1, \cdots, X_n)$ 是可交换的. ■

注释 现介绍另一种计算 $p(x_1, \cdots, x_n)$ 的方法. 将 n 个球编上号, 不妨将 k 个特殊的球编号为 $1, 2, \cdots, k$, 将 $n-k$ 个普通球编号为 $k+1, \cdots, n$. 将球一个一个地从坛子里取出来, 等价于将这些球排成一个队 $i_1, \cdots, i_n$, 第一次取出的是 i_1, 第二次取出的是 i_2, 等等, 显然 $(i_1, \cdots, i_n)$ 是一个试验结果. 而每一种排列 $(i_1, \cdots, i_n)$ 的可能性都是相等的. 而对于固定的 $(x_1, x_2, \cdots, x_n)$, 事件 $(x_1, \cdots, x_n)$ 由这样的一些试验结果 $(i_1, \cdots, i_n)$ 组成, $(i_1, \cdots, i_n)$ 中 $1, 2, \cdots, k$ 的位置与 $(x_1, \cdots, x_n)$ 中, $x_i = 1$ 所占的位置相同. 显然事件 $(x_1, \cdots, x_n)$ 中, 含有 $(i_1, \cdots, i_n)$ 这样的试验结果数为 $k\,!\,(n-k)\,!$, 这样

$$p(x_1, \cdots, x_n) = \frac{k\,!\,(n-k)\,!}{n\,!}$$

■

若 $X_1, X_2, \cdots, X_n$ 为可交换的, 很容易得到每一个 X_i 具有相同的分布. 例如,

X 和 Y 是可交换的离散随机变量, 则

$$P\{X=x\}=\sum_y P\{X=x,Y=y\}=\sum_y P\{X=y,Y=x\}=P\{Y=x\}$$

在例 8a 中, 第 i 次抽得的球为特殊球的概率等于 $P\{X_i=1\}=P\{X_1=1\}=k/n$. 直观上也是很清楚的, 坛中的 n 个球中的任意一个球在第 i 次被抽到的可能性都是一样的, 因此, $P\{X_i=1\}=k/n$.

例 8b 在例 8a 中令 Y_1 表示抽到的第一个特殊球所需的抽球个数. 令 Y_2 表示抽到第一个特殊球以后, 直到抽到第二个特殊球所需要的附加抽球的次数. 一般情况下, 令 Y_i 表示抽到 $i-1$ 个特殊球以后, 直到抽到第 i 个特殊球所需附加抽球次数, $i=1,\cdots,k$. 例如, 如果 $n=4,k=2,X_1=1,X_2=0,X_3=0,X_4=1$, 则此时, $Y_1=1,Y_2=3$. $Y_1,\cdots,Y_k$ 与 $X_1,\cdots,X_n$ 具有如下关系:

$$Y_1=i_1,Y_2=i_2,\cdots,Y_k=i_k \Leftrightarrow X_{i_1}=X_{i_1+i_2}=\cdots=X_{i_1+\cdots+i_k}=1,$$

其他的

$$X_j=0$$

由上式结合例 8a 的结论可知

$$P\{Y_1=i_1,Y_2=i_2,\cdots,Y_k=i_k\}=\frac{k\,!\,(n-k)\,!}{n\,!} \qquad i_1+\cdots+i_k\leqslant n$$

由上式看出, $Y_1,\cdots,Y_k$ 为可交换的. 现在把一副扑克牌中“A”称为特殊的牌, Y_1 表示一副洗好的扑克牌中一张一张地发牌, 直到第一张“A”出现为止所发的牌数. Y_2 表示第一张“A”以后直到第二张“A”出现为止所发的附加牌的张数, 等等. 由于 Y_1,Y_2,Y_3,Y_4 是可交换的, 因此, 所有 Y_i 的分布都相同. ■

例 8c 下面的模型称为波利亚瓮模型. 设有一个瓮含有 n 个红球, m 个蓝球. 每次从瓮中随机地抽取一个球, 记下其颜色并放回瓮中, 同时还往瓮里添加一个同颜色的球. 记

$$X_i=\begin{cases}1 & \text{如果第 } j \text{ 次抽得红球}\\ 0 & \text{如果第 } j \text{ 次抽得蓝球}\end{cases}$$

为了对 X_i 有一直观了解, 看下面的两个特殊情况

$$\begin{aligned}&P\{X_1=1,X_2=1,X_3=0,X_4=1,X_5=0\}\\&=\frac{n}{n+m}\frac{n+1}{n+m+1}\frac{m}{n+m+2}\frac{n+2}{n+m+3}\frac{m+1}{n+m+4}\\&=\frac{n(n+1)(n+2)m(m+1)}{(n+m)(n+m+1)(n+m+2)(n+m+3)(n+m+4)}\end{aligned}$$

及

$$\begin{aligned}&P\{X_1=0,X_2=1,X_3=0,X_4=1,X_5=1\}\\&=\frac{m}{n+m}\frac{n}{n+m+1}\frac{m+1}{n+m+2}\frac{n+1}{n+m+3}\frac{n+2}{n+m+4}\end{aligned}$$

$$= \frac{n(n+1)(n+2)m(m+1)}{(n+m)(n+m+1)(n+m+2)(n+m+3)(n+m+4)}$$

一般情况, 对任意 $x_1, x_2, \cdots, x_k$, 其中 r 个 1, $k-r$ 个 0,

$$P\{X_1 = x_1, \cdots, X_k = x_k\}$$
$$= \frac{n(n+1)\cdots(n+r-1)m(m+1)\cdots(m+k-r-1)}{(n+m)\cdots(n+m+k-1)}$$

由此可知, 对任意 k, 随机变量 $X_1, \cdots, X_k$ 是可交换的.

一个十分有趣的结论是, 第 i 次抽出一个红球的概率是与第一次抽取的一个红球的概率是相同的 [等于 $n/(n+m)$]. 我们可以这样直观地看这个原来不很直观的问题, 设想坛子里原来一共有 $n+m$ 类球, 红 $_1$, 红 $_2$, $\cdots$, 红 $_n$, 蓝 $_1$, $\cdots$, 蓝 $_m$. 第一次抽出的球, 这 $n+m$ 类球被抽中的概率是一样的. 由于这 $n+m$ 类球完全是对称的, 第二次抽出这 $n+m$ 类中的任意一类的概率是相同的 (放回去的另外一个球的类别与抽出来的类别被认为是相同的), 这样第 j 次抽出是红球的可能性为 $n/(n+m)$. ■

最后一个是关于可交换连续变量的例子.

例 8d 设 $X_1, \cdots, X_n$ 为相互独立的 $(0,1)$ 均匀随机数, 记 $X_{(1)}, \cdots, X_{(n)}$ 为它们的次序统计量. 令

$$Y_1 = X_{(1)} \qquad Y_i = X_{(i)} - X_{(i-1)} \qquad i = 2, \cdots, n$$

指出 $Y_1, \cdots, Y_n$ 为可交换的.

解: 考虑 n 维空间中的变换

$$y_1 = x_1, y_2 = x_2 - x_1, \cdots, y_n = x_n - x_{n-1}$$

其反变换为

$$x_i = y_1 + \cdots + y_i \qquad i = 1, \cdots, n$$

不难看出这个变换的雅可比行列式为 1, 因此, 利用 (7.3) 式, $Y_1, \cdots, Y_n$ 的分布密度为

$$f_{Y_1 \cdots Y_n}(y_1, \cdots, y_n) = f(y_1, y_1 + y_2, \cdots, y_1 + \cdots + y_n)$$

此处 f 是 $X_{(1)}, \cdots, X_{(n)}$ 的联合密度 (而不是 $X_1, \cdots, X_n$ 的联合密度), 因此

$$f_{Y_1, \cdots, Y_n}(y_1, \cdots, y_n) = n! \qquad 0 < y_1 < y_1 + y_2 < \cdots < y_1 + \cdots + y_n < 1$$

或等价地,

$$f_{Y_1, \cdots, Y_n}(y_1, \cdots, y_n) = n! \qquad 0 < y_i < 1, i = 1, \cdots, n, \ y_1 + \cdots + y_n < 1$$

从上述表达式看出 $f_{Y_1, \cdots, Y_n}$ 是 $y_1, \cdots, y_n$ 的对称函数, 即 $Y_1, \cdots, Y_n$ 为可交换的随机变量. ■

小 结

X 和 Y 的联合分布函数定义如下:

$$F(x,y)=P\{X\leqslant x,Y\leqslant y\}\qquad -\infty<x,y<\infty$$

所有关于 X,Y 的概率都可以由 F 得到. 为了求 X 和 Y 各自的分布函数, 利用

$$F_X(x)=\lim_{y\to\infty}F(x,y)\qquad F_Y(y)=\lim_{x\to\infty}F(x,y)$$

若 X 和 Y 均为离散型随机变量, 则它们的联合分布也是离散的, 其**联合分布列**为:

$$p(i,j)=P\{X=i,Y=j\}$$

X,Y 的各自分布列为

$$P\{X=i\}=\sum_j p(i,j)\qquad P\{Y=j\}=\sum_i p(i,j)$$

随机变量 X 和 Y 称为**联合连续的**, 如果存在一个二元函数, 称为**联合密度函数**$f(x,y)$, 使得对任意二维集合 C,

$$P\{(X,Y)\in C\}=\iint\limits_C f(x,y)\,\mathrm{d}x\,\mathrm{d}y$$

从上式可知

$$P\{x<X<x+\mathrm{d}x,y<Y<y+\mathrm{d}y\}\approx f(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

若 X 和 Y 联合连续, 则它们各自都为连续型的, 且密度函数分别为

$$f_X(x)=\int_{-\infty}^{\infty}f(x,y)\,\mathrm{d}y\qquad f_Y(y)=\int_{-\infty}^{\infty}f(x,y)\,\mathrm{d}x$$

随机变量 X 和 Y 称为**独立的**, 如果对任意集合 A 和 B, 有

$$P\{X\in A,Y\in B\}=P\{X\in A\}P\{Y\in B\}$$

若联合分布函数 (或者离散情形下的联合分布列, 或者连续情形下的联合密度) 可以分解为两个因子, 其中一个只依赖于 x, 另一个只依赖于 y, 则 X 和 Y 独立.

一般情况下, 随机变量 $X_1,\cdots,X_n$ 称为相互独立的, 若对一切实数集 $A_1,\cdots,$ A_n, 有

$$P\{X_1\in A_1,\cdots,A_n\in A_n\}=P\{X_1\in A_1\}\cdots P\{X_n\in A_n\}$$

若 X 和 Y 为独立连续型随机变量, 则它们的和的分布函数可以通过下式得到:

$$F_{X+Y}(a)=\int_{-\infty}^{\infty}F_X(a-y)f_Y(y)\,\mathrm{d}y$$

若 $X_i,i=1,\cdots,n$, 为独立正态随机变量, 参数分别为 μ_i 和 $\sigma_i^2,i=1,\cdots,n$, 则 $\sum_{i=1}^n X_i$ 也为正态分布随机变量, 参数为 $\sum_{i=1}^n\mu_i$ 和 $\sum_{i=1}^n\sigma_i^2$.

若 $X_i, i=1,\cdots,n$, 为独立泊松随机变量, 参数分别为 $\lambda_i, i=1,\cdots,n$, 则 $\sum_{i=1}^n X_i$ 也服从泊松分布, 参数为 $\sum_{i=1}^n \lambda_i$.

若 X 和 Y 为离散型随机变量, 则已知 $Y=y$ 的条件下, $X=x$ 的条件分布列如下定义:

$$P\{X=x|Y=y\}=\frac{p(x,y)}{p_Y(y)}$$

其中 $p(x,y)$ 为 X,Y 的联合分布列. 若 X,Y 的联合连续且其相应的联合密度为 f, 则 X 在给定 $Y=y$ 之下的条件密度函数为

$$f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}$$

设 $X_1,\cdots,X_n$ 为独立同分布的随机变量序列, 将它们进行排序以后得到 $X_{(1)}\leqslant X_{(2)}\leqslant\cdots\leqslant X_{(n)}$ 称为 $X_1,\cdots,X_n$ 的次序统计量. 设这些随机变量是连续的, 即存在密度函数 $f(x_i)$, 则它们的次序统计量的密度函数为

$$f(x_1,\cdots,x_n)=n!\,f(x_1)\cdots f(x_n)\qquad x_1\leqslant x_2\leqslant\cdots\leqslant x_n$$

随机变量序列 $X_1,\cdots,X_n$ 称为可交换的, 若对每一个 $\{1,2,\cdots,n\}$ 的排列 $i_1,\cdots,i_n$, 其相应的 $X_{i_1},\cdots,X_{i_n}$ 的联合分布是相同的.

习　题

1. 掷两枚均匀骰子, 求以下随机变量 X 和 Y 的联合分布列:

(a) X 为两枚骰子点数的最大值, Y 为两枚骰子点数之和;

(b) X 为第一枚骰子的点数, Y 为两枚骰子点数的最大值;

(c) X 为两枚骰子点数的最小值, Y 为两枚骰子点数的最大值.

2. 坛子里有 5 个白球和 8 个红球, 从中无放回地随机取出 3 个球. 如果第 i 次取出的球是白色的, 令 X_i 等于 1, 否则等于 0. 求以下联合分布列: (a) X_1,X_2; (b) X_1,X_2,X_3.

3. 在 习题 2 中, 假设白球标有数字号码, 如果第 i 个白球被取出, 那么令 Y_i 等于 1, 否则等于 0. 求以下联合分布列: (a) Y_1,Y_2; (b) Y_1,Y_2,Y_3.

4. 在有放回情形下重做习题 2.

5. 在有放回情形下重做习题 3(a).

6. 已知 5 个元件里有 2 个是失效的. 现进行检测, 每次随机地从未检测的元件中取出一个进行检测, 直到失效的都查出来为止. 记 N_1 表示第一个失效元件被检测出时的检测次数, 记 N_2 表示其后直到第二个失效元件被检测出时的附加检测次数. 求 N_1 和 N_2 的联合分布列.

7. 考虑一列独立伯努利试验, 每次成功的概率为 p. 令 X_1 表示第一次成功前失败的次数, 令 X_2 表示前两次成功之间失败的次数, 求 X_1 和 X_2 的联合分布列.

8. 设 X 和 Y 的联合密度函数如下：

$$f(x,y)=c(y^2-x^2)\mathrm{e}^{-y} \qquad -y\leqslant x\leqslant y, 0<y<\infty$$

(a) 求 c 的值； (b) 求 X 和 Y 的边缘密度； (c) 求 $E[X]$.

9. 设 X 和 Y 的联合密度函数为：

$$f(x,y)=\frac{6}{7}\left(x^2+\frac{xy}{2}\right) \qquad 0<x<1, 0<y<2$$

(a) 证明上式确实是联合密度函数； (b) 求 X 的密度函数； (c) 求 $P\{X>Y\}$;

(d) 求 $P\left\{Y>\frac{1}{2}\middle|X<\frac{1}{2}\right\}$； (e) 求 $E[X]$; (f) 求 $E[Y]$.

10. X 和 Y 的联合密度函数如下：

$$f(x,y)=\mathrm{e}^{-(x+y)} \qquad 0\leqslant x<\infty, 0\leqslant y<\infty$$

求 (a) $P\{X<Y\}$; (b) $P\{X<a\}$.

11. 某电器商店店主统计在进入该店的顾客中, 45% 的顾客会购买一台普通电视机, 15% 的顾客会购买一台等离子电视机, 剩下 40% 的顾客只是看看, 不会购买. 如果某天有 5 名顾客进入该店, 问该店将要卖出 2 台普通电视机和 1 台等离子电视机的概率是多大.

12. 某个小时内, 进入药店的人数为参数 $\lambda=10$ 的泊松随机变量, 求在该小时内已有 10 个女士进入该店的条件下, 至多有 3 位男士进入该店的条件概率. 其中作了何种假设?

13. 某男和某女约好下午 12 点半在某个地点见面. 如果男士到达的时间服从 12 点 15 到 12 点 45 之间的均匀分布, 而女士到达的时间服从 12 点到下午 1 点之间的均匀分布. 且两者的到达时间相互独立. 求先到达者等待时间不超过 5 分钟的概率. 又该男先到达的概率是多大?

14. 一辆救护车沿着长为 L 的道路来回匀速行驶, 假定某个时刻事故发生的地点服从均匀分布 [即, 事故发生地点与道路端点的距离服从 $(0,L)$ 上均匀分布]. 假设事故发生时救护车的地点也是服从均匀分布, 并与事故发生地点独立, 那么求救护车与事故发生点的距离分布.

15. 随机向量 (X,Y) 称为服从平面区域 R 上均匀分布, 若其密度函数具有下列形式

$$f(x,y)=\begin{cases} c & \text{如果}(x,y)\in R \\ 0 & \text{其他} \end{cases}$$

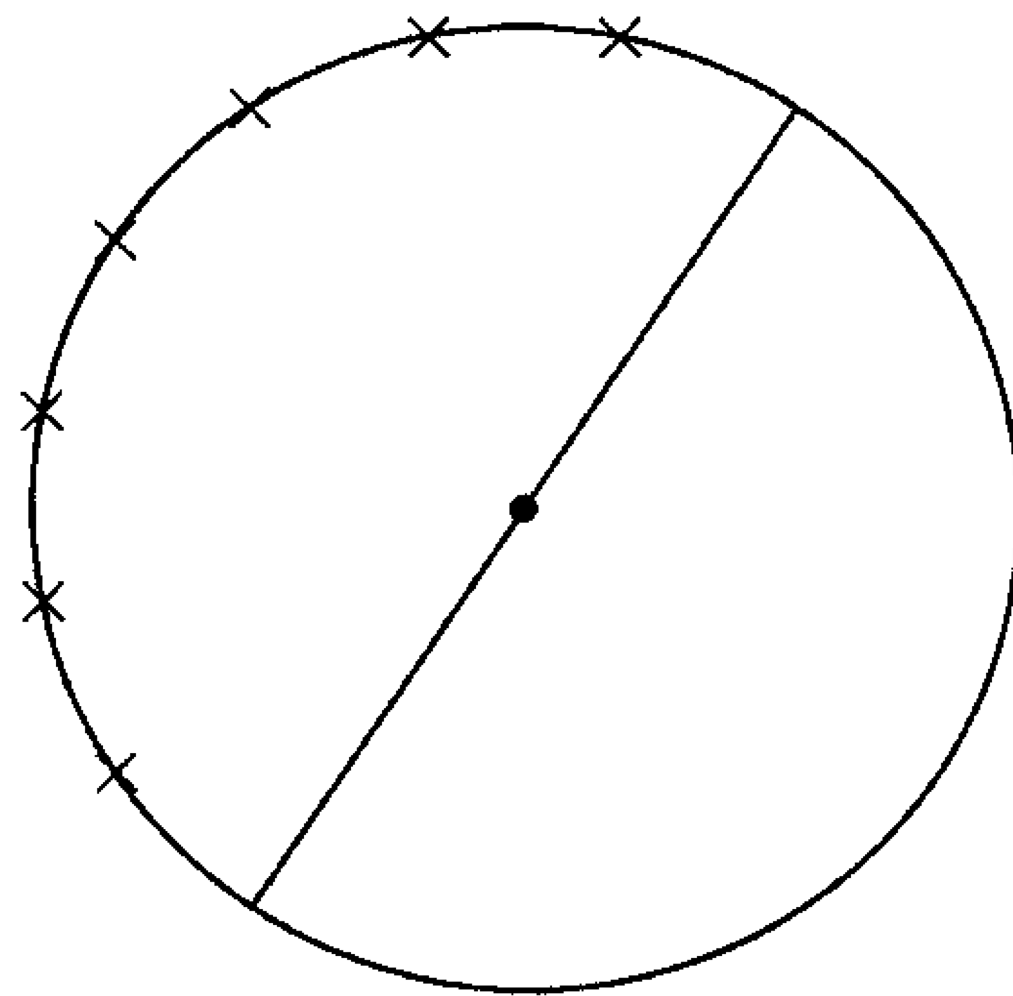
(a) 证明 $1/c$ 等于区域 R 的面积.

假定 (X,Y) 在边长为 2, 中心为 $(0,0)$ 的正方形区域内服从均匀分布.

(b) 证明 X 和 Y 相互独立, 且都服从 $(-1,1)$ 上均匀分布.

(c) 求 (X,Y) 位于以原点为圆心, 半径为 1 的圆内的概率, 也即, 求 $P\{X^2+Y^2\leqslant 1\}$ 的概率.

16. 在一个圆周上随机独立地取 n 个点, 求它们正好在一个半圆周上的概率 (也即, 我们要求可以从圆心画一根直线, 这些点恰好都在这条直线的同一边的概率). 令 $P_1, \cdots, P_n$ 表示这 n 个点, 令 A 表示事件“所有的点恰好位于某个半圆上”, 令 A_i 表示事件“处于以 P_i 为起点, 顺时针方向的一个半圆周上”.



(a) 将事件 A 用事件 $A_i, i = 1, \cdots, n$ 表示出来;

(b) A_i 之间是否互不相容?

(c) 求 $P(A)$.

17. 从一条直线 L 上随机取 X_1, X_2, X_3 三个点, 求 X_2 在 X_1 和 X_3 之间的概率.

18. 在一条长为 L 的线段的中点两边分别随机取一个点 (也即两个点 X 和 Y 为相互独立随机变量, 且 X 在 $(0, L/2)$ 上均匀分布, 而 Y 在 $(L/2, L)$ 上均匀分布), 求这两点之间距离大于 $L/3$ 的概率.

19. 证明 $f(x,y) = 1/x, 0 < y < x < 1$ 为一联合密度函数. 假设 f 为 X 和 Y 的联合密度函数, 求

(a) Y 的边缘密度; (b) X 的边缘密度; (c) $E[X]$; (d) $E[Y]$.

20. X 和 Y 的联合密度函数如下:

$$f(x,y) = \begin{cases} x\mathrm{e}^{-(x+y)} & x > 0, y > 0 \\ 0 & \text{其他} \end{cases}$$

问 X 和 Y 是否独立? 如果 $f(x,y)$ 由下式给出呢?

$$f(x,y) = \begin{cases} 2 & 0 < x < y, 0 < y < 1 \\ 0 & \text{其他} \end{cases}$$

21. 令

$$f(x,y) = \begin{cases} 24xy & 0 \leqslant x \leqslant 1, 0 \leqslant y \leqslant 1, 0 \leqslant x + y \leqslant 1 \\ 0 & \text{其他} \end{cases}$$

(a) 证明 $f(x,y)$ 为一联合密度函数; (b) 求 $E[X]$; (c) 求 $E[Y]$.

22. X 和 Y 的联合密度函数如下:

$$f(x,y) = \begin{cases} x + y & 0 < x < 1, 0 < y < 1 \\ 0 & \text{其他} \end{cases}$$

(a) X 和 Y 是否独立? (b) 求 X 的密度函数; (c) 求 $P\{X + Y < 1\}$.

23. X 和 Y 的联合密度函数如下:

$$f(x,y) = \begin{cases} 12xy(1-x) & 0 < x < 1, 0 < y < 1 \\ 0 & \text{其他} \end{cases}$$

(a) X 和 Y 是否独立? (b) 求 $E(X)$; (c) 求 $E(Y)$;

(d) 求 $\mathrm{Var}(X)$; (e) 求 $\mathrm{Var}(Y)$.

24. 考虑独立重复试验, 每次结果是 i 的概率为 p_i, $i = 0, 1, \cdots, k$, $\sum_{i=0}^{k} p_i = 1$. 令 N 表示结果第一次不等于 0 时的试验次数, 令 X 表示该结果.

(a) 求 $P\{N = n\}, n \geqslant 1$;

(b) 求 $P\{X = j\}, j = 1, \cdots, k$;

(c) 证明 $P\{N = n, X = j\} = P\{N = n\}P\{X = j\}$;

(d) 对你来说, N 独立于 X 是不是很直观?

(e) 对你来说, X 独立于 N 是不是很直观?

25. 假设 10^6 人到达某个服务站的时间为独立随机变量, 且服从 $(0, 10^6)$ 上均匀分布 (单位: 小时). 令 N 表示在第一个小时内到达的人数. 求 $P\{N = i\}$ 的近似值.

26. 假设 A, B, C 为独立随机变量, 且均服从 $(0, 1)$ 上均匀分布.

(a) 求 A, B, C 的联合分布函数.

(b) 求方程 $Ax^2 + Bx + C = 0$ 的所有根都是实数的概率.

27. 如果随机变量 X_1 和 X_2 相互独立, 且分别服从参数为 λ_1 和 λ_2 的指数分布. 求 $Z = X_1/X_2$ 的分布, 且计算 $P\{X_1 < X_2\}$.

28. 修理一辆汽车所需时间是服从指数分布的随机变量, 参数为 $\lambda = 1$.

(a) A. J. 在时刻 0 送到修理站一辆待修的车, 而 M. J. 在时刻 t 送来一辆待修的车. 问 M. J. 的车先修好的概率是多少? (假定每辆车送到修理站的时候就立刻开始修理, 并且两辆车的修理时间是相互独立的.)

(b) 现在假定 A. J. 和 M. J. 同时把车送到修理站, 而只有将 A. J. 的汽车修理完以后才开始对 M. J. 的车进行修理. 问在 $t = 2$ 以前将 M. J. 的车修好的概率有多大?

29. 某个餐馆的周销售额为一正态随机变量, 均值为 2200 元, 标准差为 230 元, 求以下事件概率:

(a) 下两周内总销售额超过 5000 元.

(b) 下三周内至少有 2 周的周销售额超过 2000 元.

你做了何种独立性假设?

30. 吉尔的保龄球得分近似服从均值为 170, 标准差为 20 的正态分布, 而杰克的得分近似服从均值为 160, 标准差为 15 的正态分布. 如果杰克和吉尔每人玩一局, 并假设他们的得分为独立正态随机变量, 求以下事件的近似概率:

(a) 杰克得分高一些.

(b) 他们两人的总得分超过 350.

31. 根据美国国家健康统计中心数据, 25.2% 的男性和 23.6% 的女性从来不吃早餐, 随机选择 200 个男士和 200 个女士的样本, 求以下事件概率的近似值:

(a) 该 400 人中至少有 110 人从不吃早餐.

(b) 不吃早餐的女性数不小于不吃早餐的男性数.

32. 某杂志的每一页上打印错误的期望数值为 0.2, 那么一篇 10 页的文章有 (a) 0 个; (b) 2 个或以上的错误的概率分别是多大? 解释理由.

33. 全球范围内, 商业航线的飞机坠毁事故每个月平均有 2.2 起, 求以下事件概率:

(a) 下个月内超过 2 起这样的事故;

(b) 下两个月内超过 4 起这样的事故;

(c) 下三个月内超过 5 起这样的事故.

并解释理由.

34. 吉有两件事情需要完成, 设这两件事情为 1 和 2. 对于每件事情, 吉试图用 1 小时完成, 但能够完成的概率分别为 $p_i(p_1 = 0.3, p_2 = 0.4)$, 若一小时内不能完成, 则需从头开始, 再用 1 小时来完成这件事情. 问: 为完成这两件事情, 共用去 12 小时 (或以上) 的概率有多大?

35. 在习题 4 中, 求以下条件下的 X_1 的条件分布列:

(a) $X_2 = 1$; (b) $X_2 = 0$.

36. 在习题 3 中, 求以下条件下的 Y_1 的条件分布列:

(a) $Y_2 = 1$; (b) $Y_2 = 0$.

37. 在习题 5 中, 求以下条件下的 Y_1 的条件分布列:

(a) $Y_2 = 1$; (b) $Y_2 = 0$.

38. 从数集 $\{1,2,3,4,5\}$ 中随机选一个数 X, 再从集合 $\{1,2,\cdots,X\}$ 中随机取第二个数, 记为 Y.

(a) 求 X 和 Y 的联合分布列;

(b) 在 $Y = i$ 的条件下求 X 的条件分布列, $i = 1,2,3,4,5$;

(c) X 和 Y 是否独立? 为什么?

39. 掷两枚骰子, 令 X 和 Y 分别表示最大点数和最小点数, 求已知 $X = i$ 的条件下 Y 的条件分布列, $i = 1,2,\cdots,6$. X 和 Y 是否独立? 为什么?

40. X 和 Y 的联合分布列如下:

$$p(1,1) = \frac{1}{8} \qquad p(1,2) = \frac{1}{4} \qquad p(2,1) = \frac{1}{8} \qquad p(2,2) = \frac{1}{2}$$

(a) 已知 $Y = i$ 的条件下求 X 的条件分布列, $i = 1,2$.

(b) X 和 Y 是否独立?

(c) 求 $P\{XY \leqslant 3\}, P\{X+Y > 2\}, P\{X/Y > 1\}$.

41. X 和 Y 联合密度函数如下:

$$f(x,y) = x\mathrm{e}^{-x(y+1)} \qquad x > 0, y > 0$$

(a) 求给定 $Y = y$ 的条件下 X 的条件密度, 以及给定 $X = x$ 的条件下 Y 的条件密度;

(b) 求 $Z = XY$ 的密度函数.

42. X 和 Y 的联合密度如下:

$$f(x,y) = c(x^2 - y^2)\mathrm{e}^{-x} \qquad 0 \leqslant x < \infty, -x \leqslant y \leqslant x$$

求给定 $X = x$ 的条件下 Y 的条件分布.

43. 某保险公司假设每个人都有一个事故参数，而且事故参数为 λ 的人每年发生事故的次数服从泊松分布，均值为 λ．他们还假设一个新投保的人的参数值为 Γ 随机变量，Γ 随机变量的分布参数为 (s, α)．如果新投保的人在第一年内发生了 n 次事故，求出他的事故参数的条件密度，并求他在下一年发生事故数的期望值．

44. 如果 X_1, X_2, X_3 为独立随机变量，且都服从 $(0,1)$ 上均匀分布，求三者中最大数大于其他两数之和的概率．

45. 一台复杂的机器只要其 5 个发动机里至少有 3 个正常工作，这台机器就正常工作．如果任何一个发动机正常工作的时间为一随机变量，其密度函数为 $f(x) = x\mathrm{e}^{-x}, x > 0$．计算这台机器正常工作时间的密度函数．

46. 设 3 辆卡车抛锚的地点随机独立地分布在长为 L 的公路上．试求任意 2 辆卡车的抛锚地点的距离均大于 d 的概率，其中 $d \leqslant L/2$．

47. 从 (0, 1) 上均匀分布中随机抽取 5 个样本值，计算其中位数落在区间 $(1/4, 3/4)$ 内的概率．

48. 设 X_1, X_2, X_3, X_4, X_5 独立同分布，且服从参数为 λ 的指数分布．求

(a) $P\{\min(X_1, X_2, \cdots, X_5) \leqslant a\}$．　(b) $P\{\max(X_1, X_2, \cdots, X_5) \leqslant a\}$．

49. 设 $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ 是 $(0,1)$ 上 n 个相互独立的均匀分布随机变量的次序统计量．求在给定 $X_{(1)} = s_1, X_{(2)} = s_2, \cdots, X_{(n-1)} = s_{n-1}$ 之下 $X_{(n)}$ 的条件分布．

50. 记 Z_1, Z_2 为相互独立的两个标准正态随机变量．令 $X = Z_1, Y = Z_1 + Z_2$，证明 X, Y 具有二元正态分布．

51. 从密度函数为 $f(x) = 2x, 0 < x < 1$ 的分布中，取容量为 2 的样本，试求出此样本的极差的分布．

52. 在以原点为圆心，半径为 1 的圆内随机取一点，令 X 和 Y 分别表示该点的两个坐标，也即其联合密度为

$$f(x,y) = \frac{1}{\pi} \qquad x^2 + y^2 \leqslant 1$$

求极坐标 $R = (X^2 + Y^2)^{1/2}$ 和 $\Theta = \arctan(Y/X)$ 的联合密度函数．

53. 设 X 和 Y 独立，且均服从 $(0,1)$ 上均匀分布．求 $R = \sqrt{X^2 + Y^2}$ 和 $\Theta = \arctan(Y/X)$ 的联合密度函数．

54. 如果 U 服从 $(0, 2\pi)$ 上均匀分布，而 Z 与 U 独立，服从参数为 1 的指数分布，直接证明 (不用例 7b 的结果) 如下定义的 X 和 Y

$$X = \sqrt{2Z}\cos U \qquad Y = \sqrt{2Z}\sin U$$

为相互独立的标准正态随机变量．

55. X 和 Y 联合密度函数为

$$f(x,y) = \frac{1}{x^2y^2} \qquad x \geqslant 1, y \geqslant 1$$

(a) 求 $U = XY$ 和 $V = X/Y$ 的联合密度函数．　(b) 求它们的边缘密度．

56. 设 X 和 Y 独立同分布，其公共分布为 $(0,1)$ 上均匀分布，求以下随机变量的联合密度：

(a) $U=X+Y, V=X/Y$; (b) $U=X, V=X/Y$; (c) $U=X+Y, V=X/(X+Y)$.

57. 在 X,Y 为独立指数分布情形下重做习题 56, 且指数分布参数为 1.

58. 设 X_1 和 X_2 独立, 且均服从参数为 λ 的指数分布, 求 $Y_1=X_1+X_2$ 和 $Y_2=\mathrm{e}^{X_1}$ 的联合密度函数.

59. 设 X,Y,Z 独立同分布, 分布密度为 $f(x)=\mathrm{e}^{-x}, 0<x<\infty$, 推导 $U=X+Y, V=X+Z, W=Y+Z$ 的联合分布.

60. 在例 8b 中, 令 $Y_{k+1}=n+1-\sum_{i=1}^{k}Y_i$, 证明 $Y_1,\cdots,Y_k,Y_{k+1}$ 可交换.

提示: 令 $Z_i=Y_i-1, i=1,\cdots,k+1$. Z_i 表示取到第 $i-1$ 个特殊球与取到第 i 个特殊球之间取到普通球的个数, $i=1,\cdots,k$. Z_{k+1} 表示取到第 k 个特殊球以后剩下的普通球的个数. 取球的次序, 刚好是将 n 个球随机地排成一列, 而 Z_i 刚好是那些特殊球之间普通球的个数. 然后证明 $Z_i, i=1,\cdots,k+1$ 的分布是可交换的.

61. 坛子里有 n 个球, 标有号码 $1,\cdots,n$. 假设从中随机取出 k 个, 令 $X_i=1$, 如果标有数字 i 的球被取出, 否则等于 0. 证明 $X_1,\cdots,X_n$ 可交换.

理 论 习 题

1. 验证 (1.2) 式.

2. 假设给定时间周期内事件发生的个数为参数 λ 的泊松分布随机变量. 事件一共分为 n 类, 每个事件属于第 i 类的概率为 $p_i, i=1,\cdots,n, \sum p_i=1$. 并且每一事件的分类不依赖于其他事件的分类. 证明发生的类型为 i 的事件数 $(i=1,\cdots,n)$ 为相互独立泊松随机变量, 参数分别为 $\lambda p_i, i=1,\cdots,n$.

3. 给出方法利用蒲丰投针问题来估算 π 值. 令人惊异的是, 这曾经一度是估算 π 值的常用方法.

4. 当 $L>D$ 时求解蒲丰投针问题.

解答: $\dfrac{2L}{\pi D}(1-\sin\theta)+2\theta/\pi$, 其中 θ 满足 $\cos\theta=D/L$.

5. 如果 X 和 Y 为独立连续正随机变量, 求以下随机变量的密度函数 (假定 X,Y 的密度函数已知):

(a) $Z=X/Y$; (b) $Z=XY$.

并在 X 和 Y 都是指数随机变量的特殊情形下求出以上的密度函数.

6. 设 X,Y 为联合连续, 其联合密度函数为 $f_{XY}(x,y)$, 证明 $X+Y$ 为连续的, 并且其密度函数为

$$f_{X+Y}(t)=\int_{-\infty}^{+\infty}f_{X,Y}(x,t-x)\mathrm{d}x$$

7. (a) 如果 X 服从参数为 (t,λ) 的 Γ 分布, 求 cX 的分布函数 $(c>0)$;

(b) 证明 $\frac{1}{2\lambda}\chi_{2n}^2$ 服从参数为 (n,λ) 的 Γ 分布, 其中 n 为正整数, χ_{2n}^2 为自由度 $2n$ 的 χ^2 分布随机变量.

8. 令 X 和 Y 为相互独立连续随机变量, 危险率函数分别为 $\lambda_X(t)$ 和 $\lambda_Y(t)$. 令 $W=\min(X,Y)$.

(a) 假定 X, Y 的分布函数已知, 求 W 的分布函数;

(b) 证明 W 的危险率函数 $\lambda_W(t)$ 由下式给出

$$\lambda_W(t) = \lambda_X(t) + \lambda_Y(t).$$

9. 令 $X_1, \cdots, X_n$ 为独立同分布随机变量, 分布为参数 λ 的指数分布. 求 $\min(X_1, \cdots, X_n)$ 的分布函数.

10. 电池的寿命为独立指数分布随机变量, 参数都为 λ. 某个手电筒工作需要 2 节电池, 如果某人有一个手电筒, 和 n 节电池, 那么该手电筒能工作的总时间的分布是什么?

11. 令 X_1, X_2, X_3, X_4, X_5 为独立同分布连续随机变量, 其公共分布函数和密度函数分别为 F 和 f. 令

$$I = P\{X_1 < X_2 < X_3 < X_4 < X_5\}$$

(a) 证明 I 与 F 无关;

提示: 将 I 写成 5 维积分, 且作变量替换: $u_i = F(x_i), i = 1, \cdots, 5$.

(b) 求 I 的值.　(c) 给出 (b) 的答案的直观解释.

12. 证明联合连续 (离散) 型随机变量 $X_1, \cdots, X_n$ 相互独立当且仅当其联合密度函数 (分布列)$f(x_1, \cdots, x_n)$ 可以写成

$$f(x_1, \cdots, x_n) = \prod_{i=1}^{n} g_i(x_i)$$

其中 $g_i(x), i = 1, \cdots, n$ 为非负函数.

13. 在例 5d[①] 中, 试验成功的概率被看成随机变量. 在该例中, 我们计算了在 $n + m$ 次试验中有 n 次成功的条件下, 试验成功概率的条件密度. 如果我们指明了哪 n 次试验成功, 相应的条件密度会不会改变?

14. 假设 X 和 Y 为独立同分布几何随机变量, 参数为 p.

(a) 不用任何计算, 你认为以下的值是多少?

$$P\{X = i | X + Y = n\}$$

提示: 假设连续掷一枚硬币, 正面朝上的概率为 p. 如果第二次正面朝上发生在第 n 次投掷, 那么第一次正面朝上的时刻的分布列是什么?

(b) 验证 (a) 的猜测.

15. 设一独立试验序列中, 每次试验成功的概率为 p. 已知第 n 次试验的结果为第 k 次成功. 指出前 $n - 1$ 次试验的各种可能的结果是等概的 (每一个结果为 $k - 1$ 次成功和 $n - k$ 次失败组成的一个序列).

16. 如果 X 和 Y 为独立同分布二项随机变量, 参数为 (n, p), 用分析的方法证明, 给定 $X+Y = m$ 的条件下 X 的条件分布为超几何分布. 另外, 不用任何计算给出此结果的另一种解释.
提示: 考虑掷 $2n$ 枚硬币, 令 X 表示前 n 个硬币出现正面朝上的枚数, Y 表示后 n 枚硬币出现的正面数. 说明在给定有 m 次正面朝上的情况下, 前 n 枚硬币出现的正面数与从 n 个白球 n 个黑球里随机抽取 m 个球, 其中的白球数是同分布的.

---

① 原文写成了例 5c, 有误. —— 译者注

17. 设 $X_i(i=1,2,3)$ 为独立泊松随机变量, 参数分别为 $\lambda_i, i=1,2,3$. 令 $X=X_1+X_2$ 且 $Y=X_2+X_3$, 随机向量 (X, Y) 称为服从二元泊松分布. 求其联合分布列, 即求 $P\{X=n, Y=m\}$.

18. 设 X 和 Y 均为整数值随机变量. 记

$$p(i|j)=P(X=i|Y=j), \quad q(j|i)=P(Y=j|X=i)$$

证明

$$P(X=i, Y=j)=\frac{p(i|j)}{\sum\limits_i \frac{p(i|j)}{q(j|i)}}$$

19. 令 X_1, X_2, X_3 为独立同分布连续随机变量, 计算

(a) $P\{X_1>X_2|X_1>X_3\}$; (b) $P\{X_1>X_2|X_1<X_3\}$;

(c) $P\{X_1>X_2|X_2>X_3\}$; (d) $P\{X_1>X_2|X_2<X_3\}$.

20. 令 U 表示 $(0,1)$ 上均匀随机变量, 计算下列条件下的 U 的条件分布:

(a) $U>a$; (b) $U<a$, 其中 $0<a<1$.

21. 假设给定某天内, 空气中的水分含量 W 为参数 (t,β) 的 Γ 随机变量, 也即, 其密度为 $f(w)=\beta \mathrm{e}^{-\beta w}(\beta w)^{t-1}/\Gamma(t), w>0$. 给定 $W=w$, 假定当天内事故发生次数 N 服从参数为 w 的泊松分布. 证明给定 $N=n$ 的条件下, W 的条件分布为参数 $(t+n, \beta+1)$ 的 Γ 分布.

22. 令 W 为参数 (t,β) 的 Γ 随机变量, 并假设在 $W=w$ 的条件下, $X_1, X_2, \cdots, X_n$ 为独立指数型随机变量, 参数为 w. 证明, 给定 $X_1=x_1, X_2=x_2, \cdots, X_n=x_n$ 条件下, W 的条件分布为 Γ 分布, 参数为 $\left(t+n, \beta+\sum_{i=1}^n x_i\right)$.

23. mn 个数的矩阵排成了 n 行, 每行 m 列, 如果有一个数, 既是该行的最小值, 又是该列的最大值, 那么就称它为鞍点(saddlepoint). 比如, 在以下的方阵中

$$\begin{matrix} 1 & 3 & 2 \\ 0 & -2 & 6 \\ 0.5 & 12 & 3 \end{matrix}$$

第 1 行第 1 列里的 1 便是一个鞍点. 鞍点的存在性在博弈论里是一个很重要的问题. 考虑一个如上所述的关于数的矩阵, 并假设有两个人 A 和 B 进行如下游戏: A 从 $1,2,\cdots,n$ 中抽取一数, B 从 $1,2,\cdots,m$ 中抽取一数. 并规定它们同时宣布自己的选择, 如果 A 抽到了 i 而 B 抽到了 j, 那么 B 将付给 A 的钱数为该数阵里的第 i 行第 j 列的数. 假设该阵有一个鞍点, 不妨假设为第 r 行第 k 列, 该数为 x_{rk}. 这样, 如果 A 抽取了 r, 那么他可保证至少赢得 x_{rk}(因为 x_{rk} 为第 r 行里的最小数, 如果他选择其他行, 就不能保证至少赢得 x_{rk}.), 另一方面, 如果 B 抽取了 k, 那么他可保证至多输掉 x_{rk}(因为 x_{rk} 为 k 列中的最大数, 如果 B 选择其他列, 他输的钱可能更多.). 由于 A 有方法可以保证他至少赢得 x_{rk}, 而 B 有方法可保证他至多输掉 x_{rk}, 因此, 该策略对两人来说是最佳的.

如果该矩阵里的 nm 个数是从任意一个连续分布里独立抽取的, 那么该阵具有鞍点的概率是多大?

24. 如果 X 服从指数分布, 参数为 λ, 计算 $P\{[X]=n, X-[X]\leqslant x\}$, 其中记号 $[x]$ 定义为不超过 x 的最大整数. 你能推导出 $[X]$ 同 $X-[X]$ 是独立的吗?

25. 假设 $F(x)$ 为一分布函数, 证明当 n 为整数时, 以下都是分布函数:

(a) $F^n(x)$;　(b) $1-[1-F(x)]^n$.

提示: 令 $X_1,\cdots,X_n$ 为独立同分布随机变量, 其分布函数为 F. 用 X_i 来定义随机变量 Y 和 Z, 以满足 $P\{Y\leqslant x\}=F^n(x)$, 且 $P\{Z\leqslant x\}=1-[1-F(x)]^n$.

26. 证明: 如果 n 个人随机分布在长度为 L 英里的路上, 那么两两之间的最小距离大于 D 英里的概率为 $[1-(n-1)D/L]^n$, 其中 $D\leqslant L/(n-1)$. 如果 $D>L/(n-1)$ 呢?

27. 通过对公式 (6.4) 求微分, 建立公式 (6.2).

28. 证明: 取自 $(0,1)$ 均匀分布的 $2n+1$ 个相互独立的、样本的中位数服从 β 分布, 参数为 $(n+1,n+1)$.

29. 验证公式 (6.6) 是 $X_{(i)}$ 和 $X_{(j)}$ 的联合密度.

30. 求密度函数为 f 的连续分布的 n 个样本的极差的密度.

31. 令 $X_{(1)}\leqslant X_{(2)}\leqslant\cdots\leqslant X_{(n)}$ 为 n 个相互独立的 $(0,1)$ 上均匀随机变量的次序统计量. 证明, 对 $1\leqslant k\leqslant n+1$, 有

$$P\{X_{(k)}-X_{(k-1)}>t\}=(1-t)^n$$

其中 $X_{(0)}\equiv 0, X_{(n+1)}\equiv 1$①.

32. 令 $X_1,\cdots,X_n$ 为一列独立同分布连续随机变量, 分布函数为 F. 又令 $X_{(i)}, i=1,\cdots,n$ 表示相应的次序统计量, 如果 X 独立于 $X_i, i=1,\cdots,n$, 并且也具有分布 F. 计算

(a) $P\{X>X_{(n)}\}$;　(b) $P\{X>X_{(1)}\}$;　(c) $P\{X_{(i)}<X<X_{(j)}\}, 1\leqslant i<j\leqslant n$.

33. 令 $X_1,\cdots,X_n$ 为独立同分布连续随机变量, 分布函数为 F, 密度函数为 f. 随机变量 $M\equiv[X_{(1)}+X_{(n)}]/2$ 定义为其最大值和最小值的平均值, 称为中程. 证明它的分布函数为

$$F_M(m)=n\int_{-\infty}^{m}[F(2m-x)-F(x)]^{n-1}f(x)\,\mathrm{d}x$$

34. 令 $X_1,\cdots,X_n$ 为 (0, 1) 上独立均匀随机变量. 令 $R=X_{(n)}-X_{(1)}$ 表示极差, $M=[X_{(n)}+X_{(1)}]/2$ 表示中程, 计算 R 和 M 的联合密度.

35. 令 X 和 Y 为相互独立的标准正态随机变量, 计算以下随机变量的联合密度:

$$U=X\qquad V=\frac{X}{Y}$$

然后利用此结果证明 X/Y 具有柯西分布.

自 检 习 题

1. 掷一枚非均匀骰子, 每个奇数面 1, 3, 5 朝上的概率为 C, 每个偶数面朝上的概率为 $2C$.

(a) 求 C 的值.

① 原文误写成 $X_{(n+1)}\equiv t$. —— 译者注

(b) 投掷这枚骰子, 令

$$X=\begin{cases}1 & \text{结果为偶数}\\ 0 & \text{其他}\end{cases} \quad \text{及} \quad Y=\begin{cases}1 & \text{结果大于 } 3\\ 0 & \text{其他}\end{cases}$$

求 X 和 Y 的联合分布列.

现在假设独立投掷 12 次骰子.

(c) 求 6 个面中的每一面正好出现 2 次的概率.

(d) 求其中点数为 1 或 2 的有 4 次, 点数为 3 或 4 的有 4 次, 点数为 5 或 6 的有 4 次的概率.

(e) 求至少有 8 次点数为偶数的概率.

2. 随机变量 X, Y, Z 的联合分布列如下:

$$p(1,2,3)=p(2,1,1)=p(2,2,1)=p(2,3,2)=\frac{1}{4}$$

求 (a) $E[XYZ]$. (b)$E[XY+XZ+YZ]$.

3. X 和 Y 的联合密度如下

$$f(x,y)=C(y-x)\mathrm{e}^{-y} \qquad -y<x<y, 0<y<\infty$$

(a) 求 C; (b) 求 X 的密度函数; (c) 求 Y 的密度函数;

(d) 求 $E[X]$; (e) 求 $E[Y]$.

4. 记 $r=r_1+\cdots+r_k$, 其中 r_i 为正整数. 证明, 若 $X_1, \cdots, X_r$ 具有多项分布, 则 $Y_1, \ldots, Y_k$ 也具有多项分布, 其中

$$Y_i=\sum_{j=r_{i-1}+1}^{r_{i-1}+r_i} X_j, \quad i \leqslant k$$

上式中 r_0 约定为 0, 如此定义的 Y_i 表示, Y_1 是前面 r_1 个 X_j 的和, Y_2 是接下来 r_2 个 X_j 之和, 等等.

5. 设 X, Y, Z 为独立随机变量, 每个都等可能地取值 1 或 2, 求以下随机变量的概率分布列:

(a) XYZ; (b) $XY+XZ+YZ$; (c) X^2+YZ.

6. 设 X 和 Y 为连续型随机变量, 其联合密度为

$$f(x,y)=\begin{cases}\dfrac{x}{5}+cy & 0<x<1, 1<y<5\\ 0 & \text{其他}\end{cases}$$

其中 c 为一常数.

(a) c 的值是多少? (b) X 和 Y 是否独立? (c) 求 $P\{X+Y>3\}$.

7. X 和 Y 的联合密度为

$$f(x,y)=\begin{cases}xy & 0<x<1, 0<y<2\\ 0 & \text{其他}\end{cases}$$

(a) X 和 Y 是否独立? (b) 求 X 的密度函数; (c) 求 Y 的密度函数;

(d) 求联合分布函数;　(e) 求 $E[Y]$;　(f) 求 $P\{X+Y<1\}$.

8. 假设有 2 个元件和 3 种撞击, 其中第一种撞击将引起第一个元件失效, 第二种撞击引起元件 2 失效, 第三种撞击将导致两个元件都失效. 等待三种撞击的时间为独立指数随机变量, 参数分别为 λ_1, λ_2 和 λ_3. 令 X_i 表示元件 $i, i=1,2$ 失效的时刻, 随机变量 X_1, X_2 称为服从联合二元指数分布. 计算 $P\{X_1>s, X_2>t\}$.

9. 假设一分类广告册有 m 页, 其中 m 充分大. 假设每页的广告数量是个变量, 而且你要知道某页上有多少广告唯一的方法就是直接计数. 还有, 假设页数足够多, 你无法直接数出到底有多少广告. 你的目标是想随机选择一个广告以使得任何一个广告被选中的可能性是一样的.

(a) 随机挑选一页, 然后随机从中选择一个广告. 这样是否达到了你的目标? 为什么?

令 $n(i)$ 表示第 i 页上的广告数 $i=1,\cdots,m$. 尽管这些数 $n(i)$ 是未知的, 但我们仍假定它们都小于或等于某个给定的值 n. 考虑如下选择广告的法则:

第 1 步　随机选择一页, 假设页数为 X, 直接数出该页上的广告数以得到 $n(X)$;

第 2 步　以概率 $n(X)/n$ 决定是否继续考察第 X 页, 如果决定继续考察第 X 页, 转向第 3 步, 否则回到第 1 步;

第 3 步　在第 X 页上随机选择一个广告.

每经过"第 1 步"一次称为一次循环. 比如, 第 1 步随机选择的页码被拒绝继续考察, 而第二次选择的页码接受了, 即转入第 3 步, 那么我们需要两次循环才能得到一个广告.

(b) 一次循环就能得到第 i 页上广告的概率是多大?

(c) 一次循环就能选中一个广告的概率是多大?

(d) 通过 k 次循环才选中了第 i 页上的第 j 个广告的概率是多大?

(e) 通过法则, 第 i 页上的第 j 个广告被选中的概率是多大?

(f) 通过法则, 循环的次数的期望值是多大?

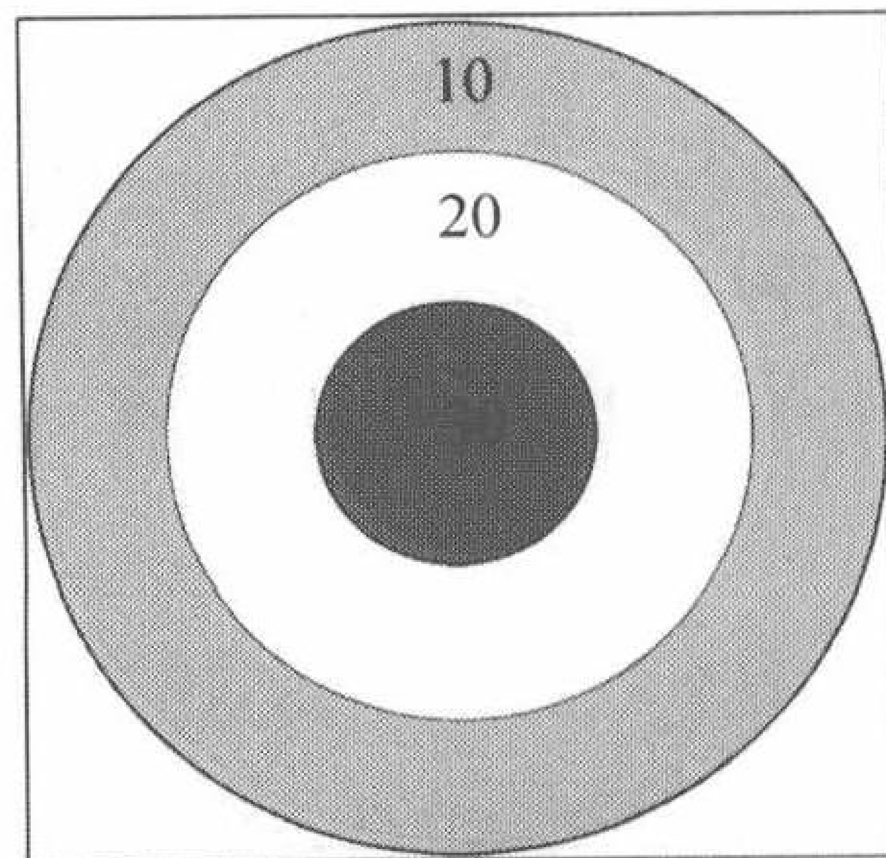
10. 在自检习题 8 中的"随机"部分可通过 $(0,1)$ 均匀随机变量实现.

第 1 步　产生一 $(0,1)$ 均匀随机数 U, 令 $X=[mU]+1$, 找到 X 页上的广告数 $n(X)$. (其中 $[x]$ 表示不超过 x 的最大整数, 或 x 的整数部分.)

(a) 解释为什么上述算法实现了自检习题 8 中的第 1 步?

提示: X 的分布列是什么?

(b) 写出实现自检习题 8 中其余各步的算法.



11. 令 $X_1, X_2, \cdots, X_n$ 为一列 (0, 1) 上独立均匀随机变量, 对于一个给定常数 c, 定义如下随机变量: $N=\min\{n: X_n>c\}$, N 是否独立于 X_N? 即如果知道第一个大于 c 的随机变量的值, 是否影响这个随机变量发生时刻的概率分布? 给出您的答案的直观解释.

12. 右图中的镖靶是个边长为 6 的正方形. 中间有三个半径分别为 1, 2, 3 的同心圆圈. 如果镖落入半径为 1 的圆内, 那么得分为 30. 如果落入半径为 1 的圆外, 但是

落入半径为 2 的圆内, 得分 20. 如果落入半径为 2 的圆外, 但是落入半径为 3 的圆内, 得分为 10. 其他情况得分为 0. 假设你每次掷镖与前面的结果都是独立的, 落点均匀分布在该正方形内, 求以下事件概率:

(a) 一次投镖得分 20; (b) 一次投镖得分至少为 20;

(c) 一次投镖得分为 0; (d) 一次掷镖的期望得分值;

(e) 两次掷镖每次得分至少为 10; (f) 两次掷镖后总得分为 30.

13. NBA 篮球赛中有这样的规律: 两支实力相当的球队比赛的时候, 每节主队得分与客队得分之差为正态随机变量, 均值为 1.5, 方差为 6. 并且, 假设 4 节的比分差是相互独立的.

(a) 主队胜的概率为多大?

(b) 在前半场主队落后 5 分的情况下, 主队得胜的概率为多大?

(c) 在第一节主队赢 5 分的情况下, 主队得胜的概率为多大?

14. 令 N 为几何随机变量, 参数为 p. 假设已知 $N = n$ 的条件下 X 的条件分布为参数为 (n, λ) 的 Γ 分布. 求已知 $X = x$ 的条件下 N 的条件分布列.

15. 令 X 和 Y 为 $(0,1)$ 上相互独立的均匀随机变量.

(a) 求 $U = X, V = X + Y$ 的联合密度. (b) 利用 (a) 得到的结果求 V 的密度函数.

16. 你和其他三个人正参与一个项目的竞拍, 竞价高者获胜. 如果你中标, 便计划立即将此项目转售, 售价为 10 000 美元. 如果你认为其他人的竞价是独立的, 且在 7000 到 11 000 美元之间均匀分布, 你如何竞价才能使得期望获益最大? (如果你中标, 你必须将此项目以你的出价买下.)

17. 设 $X_1, X_2, \cdots, X_n$ 为独立随机变量, 且

(a) 均为等可能取 $1, \cdots, n$ 中任一值. (b) 具有分布列 $P\{X_i = j\} = p_j, j = 1, 2, \cdots, n$.

分别在以上两种情形下, 求出 $X_1, X_2, \cdots, X_n$ 恰为 $1, 2, \cdots, n$ 的一个排列的概率.

18. 令 $X_1, \cdots, X_n$ 和 $Y_1, \cdots, Y_n$ 为相互独立随机向量, 其中每个向量都是 k 个 1 和 $n-k$ 个 0 的随机排列. 也即, 它们的联合概率分布列为:

$$P\{X_1 = i_1, \cdots, X_n = i_n\} = P\{Y_1 = i_1, \cdots, Y_n = i_n\} = 1\Big/\binom{n}{k} \qquad i_j = 0, 1, \sum_{j=1}^{n} i_j = k.$$

令 $N = \sum_{i=1}^{n} |X_i - Y_i|$ 表示两个向量的分量取不同值的坐标数, 而且, 令 M 表示满足 $X_i = 1, Y_i = 0$ 的 i 的个数.

(a) 求 N 和 M 的关系; (b) 求 M 的分布; (c) 求 $E[N]$; (d) 求 $\mathrm{Var}(N)$.

***19.** 令 $Z_1, Z_2, \cdots, Z_n$ 为独立标准正态随机变量, 且令 $S_j = \sum_{i=1}^{j} Z_i$.

(a) 求 $S_k = y$ 的条件下的 S_n 的条件分布, $k = 1, 2, \cdots, n$;

(b) 证明: 对于 $1 \leqslant k \leqslant n$, 已知 $S_n = x$ 的条件下, S_k 的条件分布为均值 xk/n, 方差 $k(n-k)/n$ 的正态分布.

20. 令 $X_1, X_2, \cdots$ 为一列独立同分布连续随机变量, 求

(a) $P\{X_6 > X_1 | X_1 = \max(X_1, \cdots, X_5)\}$; (b) $P\{X_6 > X_2 | X_1 = \max(X_1, \cdots, X_5)\}$.

第 7 章 期望的性质

7.1 引 言

本章中, 我们将进一步讨论期望的性质. 对于离散型随机变量 X, 它的期望由下式定义:

$$E[X]=\sum_{x}xp(x)$$

其中 $p(x)$ 是 X 的分布列. 对于连续型随机变量, X 的期望由

$$E[X]=\int_{-\infty}^{\infty}xf(x)\mathrm{d}x$$

定义, 其中 $f(x)$ 是 X 的密度函数.

由于 $E[X]$ 是随机变量 X 的所有可能取值的加权平均, $E[X]$ 的值必定介于 X 的两个极值之间. 因此, 我们有如下结论, 若

$$P\{a\leqslant X\leqslant b\}=1$$

则

$$a\leqslant E[X]\leqslant b$$

我们在离散情形下证明此结论. 由 $P\{a\leqslant X\leqslant b\}=1$ 可知对于一切 $x\notin[a,b]$, 均有 $p(x)=0$, 因此

$$E[X]=\sum_{x:p(x)>0}xp(x)\geqslant\sum_{x:p(x)>0}ap(x)=a\times\sum_{x:p(x)>0}p(x)=a$$

利用类似方法可证 $E[X]\leqslant b$. 这样, 我们在离散情形下证明了 $a\leqslant E[X]\leqslant b$. 连续型情形下的证明完全类似, 细节从略.

7.2 随机变量和的期望

第 4 章的命题 4.1 和第 5 章的命题 2.1 给出了随机变量的函数的期望值的计算公式. 此处, 我们将这个公式推广到二元函数的情况. 设 X 和 Y 为随机变量, g 是一个二元函数.

命题 2.1 若 X, Y 具有二元分布列 $p(x,y)$, 则

$$E[g(X,Y)] = \sum_y \sum_x g(x,y)p(x,y)$$

若 X, Y 具有联合分布密度 $f(x,y)$, 则

$$E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y)f(x,y)\,\mathrm{d}x\,\mathrm{d}y$$

我们在 (X,Y) 为连续情形和 $g(X,Y)$ 为非负情形下证明此命题. 由于 $g(X,Y) \geqslant 0$, 利用第 5 章的引理 2.1, 可得

$$E[g(X,Y)] = \int_0^{\infty} P\{g(X,Y) > t\}\mathrm{d}t$$

将概率

$$P\{g(X,Y) > t\} = \iint_{(x,y):g(x,y)>t} f(x,y)\,\mathrm{d}y\,\mathrm{d}x$$

代入期望公式, 得

$$E[g(X,Y)] = \int_0^{\infty} \iint_{(x,y):g(x,y)>t} f(x,y)\,\mathrm{d}y\,\mathrm{d}x\,\mathrm{d}t$$

将上述的三重积分交换积分次序, 得

$$\begin{aligned} E[g(X,Y)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{t=0}^{g(x,y)} f(x,y)\,\mathrm{d}t\,\mathrm{d}y\,\mathrm{d}x \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y)f(x,y)\,\mathrm{d}y\,\mathrm{d}x \end{aligned}$$

这样, 我们证明了当 $g(X,Y)$ 为非负随机变量时 $E[g(X,X)]$ 的计算公式. $g(X,Y)$ 为一般情形下, 可参考一维情况处理. (参见第 5 章的理论习题 2 和理论习题 3)

例 2a 设在长度为 L 的一段路 $[0,L]$ 上某一点 X 处发生了车祸. 在发生车祸的同时, 在 $[0,L]$ 的某一点 Y 处有一辆救护车. 假定 X, Y 都是均匀地分布在地段 $[0,L]$ 上, 并且相互独立, 求事故地点和救护车之间的平均距离.

解: X 和 Y 之间的平均距离就是 $E[|X-Y|]$, 由于 (X,Y) 的联合密度函数为

$$f(x,y) = \frac{1}{L^2}, \qquad 0 < x < L, 0 < y < L$$

由命题 2.1 可知,

$$E[|X-Y|] = \frac{1}{L^2}\int_0^L \int_0^L |x-y|\,\mathrm{d}y\,\mathrm{d}x$$

经计算,

$$\int_0^L |x-y|\mathrm{d}y = \int_0^x (x-y)\mathrm{d}y + \int_x^L (y-x)\mathrm{d}y$$
$$= \frac{x^2}{2} + \frac{L^2}{2} - \frac{x^2}{2} - x(L-x) = \frac{L^2}{2} + x^2 - xL$$

从而

$$E[|X-Y|] = \frac{1}{L^2}\int_0^L \left(\frac{L^2}{2} + x^2 - xL\right)\mathrm{d}x = \frac{L}{3}$$ ■

作为命题 2.1 的重要应用, 我们得到如下的结果. 设 $E[X]$ 和 $E[Y]$ 均有限, 令 $g(x,y) = X + Y$. 在 (X,Y) 为连续的情况下, 利用命题 2.1 可得

$$E[X+Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x+y)f(x,y)\,\mathrm{d}x\,\mathrm{d}y$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,y)\,\mathrm{d}y\,\mathrm{d}x + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x,y)\,\mathrm{d}x\,\mathrm{d}y$$
$$= \int_{-\infty}^{\infty} xf_X(x)\mathrm{d}x + \int_{-\infty}^{\infty} yf_Y(y)\,\mathrm{d}y = E[X] + E[Y]$$

一般情形下, 结论成立. 因此, 当 $E[X]$ 和 $E[Y]$ 均有限时,

$$E[X+Y] = E[X] + E[Y] \tag{2.1}$$

例 2b 设随机变量 X 和 Y 满足如下关系

$$X \geqslant Y$$

即对于任何试验结果, 随机变量 X 的值永远大于等于随机变量 Y 的值. 因此, 对任何试验结果, $X - Y \geqslant 0$ 永远成立, 由期望的定义可知, $E[X-Y] \geqslant 0$ 或等价地

$$E[X] \geqslant E[Y]$$ ■

由 (2.1) 式, 通过归纳法可知, 对于任何一组随机变量 $X_i, i = 1, 2, \cdots, n$, 只要他们的期望均有限, 就有下式成立:

$$E[X_1 + \cdots + X_n] = E[X_1] + \cdots + E[X_n] \tag{2.2}$$

公式 (2.2) 是一个十分有用的公式, 下面一系列例子说明了其有用之处.

例 2c (样本均值) 设 $X_1, \cdots, X_n$ 为独立同分布的随机变量序列, 其公共分布函数为 F, 期望值为 μ, 这样的序列称为来自分布 F 的一个样本. 由下式

$$\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n}$$

定义的 $\bar{X}$ 称为样本均值. 计算 $E[\bar{X}]$.

解:

$$E[\bar{X}] = E\left[\sum_{i=1}^{n} \frac{X_i}{n}\right] = \frac{1}{n}E\left[\sum_{i=1}^{n} X_i\right]$$
$$= \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \mu \qquad \text{因为} E[X_i] \equiv \mu$$

这个公式说明, 样本均值的期望值等于其分布的均值. 在统计中, 分布的期望值通常是未知的, 而样本的均值就作为 μ 的估计值. ■

例 2d (布尔不等式) 设 $A_1, \cdots, A_n$ 为 n 个随机事件, 记 $X_i, i = 1, \cdots, n$ 为这些事件的示性函数,

$$X_i = \begin{cases} 1 & A_i\text{发生} \\ 0 & \text{其他} \end{cases}$$

记 $X = \sum_{i=1}^{n} X_i$, 刚好是这一系列事件在试验中发生的次数, 令

$$Y = \begin{cases} 1 & X \geqslant 1 \\ 0 & \text{其他} \end{cases}$$

故当 $A_i, i = 1, \cdots, n$ 中至少有一个事件发生时, $Y = 1$, 否则 $Y = 0$. 由此立即可知 $X \geqslant Y$. 从而 $E[X] \geqslant E[Y]$. 由

$$E[X] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} P(A_i)$$

$$E[Y] = P\{A_i\text{中至少有一事件发生}\} = P\Big(\bigcup_{i=1}^{n} A_i\Big)$$

可知, 布尔不等式成立, 即

$$P\Big(\bigcup_{i=1}^{n} A_i\Big) \leqslant \sum_{i=1}^{n} P(A_i)$$ ■

下面的 3 个例子说明 (2.2) 式可用于二项分布, 负二项分布和超几何分布的期望公式的推导. 将现在的方法与第 4 章中的方法进行对比, 可显示出公式 (2.2) 的优越之处.

例 2e (具有二项分布的随机变量的期望公式) 设 X 的分布为二项分布, 其参数为 (n, p). 注意到 X 是 n 次独立重复试验中的成功次数, 而每次成功的概率为 p, 我们可将 X 写成

$$X = X_1 + X_2 + \cdots + X_n$$

其中

$$X_i=\begin{cases}1 & 第\ i\ 次试验成功\\ 0 & 第\ i\ 次试验失败\end{cases}$$

因此, X_i 是一个伯努利随机变量, 其期望 $E[X_i]=1\times p+0\times(1-p)=p$. 因此

$$E[X]=E[X_1]+E[X_2]+\cdots+E[X_n]=np$$ ■

例 2f (负二项随机变量的期望公式) 设在一独立重复试验序列中, 每次试验成功的概率为 p, 记 X 为试验序列中直到 r 次成功所需的试验次数. 现在希望计算 X 的期望值.

解: 可将 X 写成下列表达式

$$X=X_1+X_2+\cdots+X_r$$

其中 X_1 表示达到第一次成功所需要的试验次数, X_2 表示在试验序列中达到第二次成功所需的附加次数, X_i 表示第 $i-1$ 次成功以后, 为获得第 i 次成功所需的附加试验次数. 稍加思索, 就可得到结论, X_i 的分布是参数为 p 的几何分布. 由第 4 章的例 8b 可知 $E[X_i]=1/p, i=1,\cdots,r$. 因此

$$E[X]=E[X_1]+E[X_2]+\cdots+E[X_r]=r/p$$ ■

例 2g (超几何随机变量的期望) 设坛子内有 N 个球, 其中 m 个白球, 从中随机地取 n 个球, 求取出白球个数的期望值.

解: 记 X 为取出的白球个数, X 可以表示成: $X=X_1+\cdots+X_m$, 其中

$$X_i=\begin{cases}1 & 第\ i\ 个白球被取出\\ 0 & 其他\end{cases}$$

现在,

$$E[X_i]=P\{X_i=1\}=P\{第\ i\ 个白球被取出\}=\binom{1}{1}\binom{N-1}{n-1}\Big/\binom{N}{n}=n/N$$

因此

$$E[X]=E[X_1]+\cdots+E[X_m]=\frac{mn}{N}$$

我们也可用另一种表示方法求得此结果. 随机变量 X 可表示成

$$X=Y_1+\cdots+Y_n$$

其中

$$Y_i=\begin{cases}1 & 第\ i\ 次选出的是白球\\ 0 & 其他\end{cases}$$

此处的 i 是第 i 次抽取的意思, 而前面 X_i 中的 i 是第 i 个白球的意思, 两者含义不同. 由于 N 个球中的每一个球被第 i 次取出的概率都相同, 因此 $E[Y_i]=m/N$, 故

$$E[X]=E[Y_1]+\cdots+E[Y_n]=nm/N$$

■

例 2h (配对的期望数) N 个人首先把他们的帽子丢在某房间内, 将帽子充分混合以后, 又让每一个人随机地取一顶帽子, 求选中自己帽子的人数的期望值.

解: 记 X 为选中自己帽子的人数, X 可写成: $X=X_1+X_2+\cdots+X_N$, 其中

$$X_i=\begin{cases}1 & \text{第 } i \text{ 个人选中自己的帽子}\\ 0 & \text{其他}\end{cases}$$

对于每一个人来说, 选中任何一顶帽子的可能性是相同的, 因此

$$E[X_i]=P\{X_i=1\}=\frac{1}{N} \qquad i=1,\cdots,N$$

这样

$$E[X]=E[X_1]+\cdots+E[X_N]=\frac{1}{N}\times N=1$$

因此, 平均来说, 只有一个人能拿到自己的帽子. ■

例 2i (优惠券的收集问题) 设一共有 N 种不同的优惠券, 假定有一人在收集优惠券, 每次得到一张优惠券, 而得到的优惠券在这 N 种优惠券中均匀地分布. 求出当这人收集到全套 N 张优惠券的时候, 他收集到的优惠券张数的期望值.

解: 记 X 表示这人收集到全套优惠券时所收集的优惠券的总数. 我们利用例 2f 中的方法来计算 $E[X]$. 记 X_i 表示第 i 种优惠券已经收集到, 为收集到第 $i+1$ 种优惠券所需要的附加次数. 注意 X 具有如下表达式:

$$X=X_0+X_1+\cdots+X_{N-1}$$

设有 i 种优惠券已经收集到, 则下一次收集到一张新的优惠券的概率为 $(N-i)/N$. 这样, X_i 的分布为几何分布, 即

$$P\{X_i=k\}=\frac{N-i}{N}\left(\frac{i}{N}\right)^{k-1} \qquad k\geqslant 1$$

因此, $E[X_i]=N/(N-i)$, 由此可得

$$E[X]=1+\frac{N}{N-1}+\frac{N}{N-2}+\cdots+\frac{N}{1}=N\left[1+\frac{1}{2}+\cdots+\frac{1}{N-1}+\frac{1}{N}\right]$$

■

例 2j 10 个猎人等待一批野鸭飞过, 当一群 10 只野鸭飞过猎人头顶时, 10 个猎人随机地瞄准一只野鸭并同时射击. 设每位猎人击中野鸭的概率为 p, 求逃过这一劫的野鸭数的期望值.

解: 记

$$X_i = \begin{cases} 1 & \text{第 } i \text{ 只野鸭逃过这一劫} \\ 0 & \text{其他} \end{cases}$$

于是,

$$E[X] = E[X_1 + \cdots + X_{10}] = E[X_1] + \cdots + E[X_{10}] = 10E[X_1]$$

其中, X 表示逃过这一劫的野鸭数, $E[X_i] = P\{X_i = 1\}$ 表示 1 号野鸭未被击中的概率. 每位猎手是否击中 i 号野鸭是相互独立的, 且概率为 $p/10$, 因此, $P\{X_i = 1\} = (1 - p/10)^{10}$, 从而 $E[X] = 10(1 - p/10)^{10}$. ■

例 2k (游程的期望数) 设有 n 个 1 和 m 个 0 随机地排成一个序列, 一共有 $(n+m)!/(n!m!)$ 种不同的排列法. 假定每种排列法都是等可能的. 在一个排列中, 连在一起的 1 构成"1"的游程. 例如, $n = 6, m = 4$, 6 个 1 和 4 个 0 构成如下的一个排列: 1, 1, 1, 0, 1, 1, 0, 0, 1, 0, 其中第一组 3 个 1 构成一个"1"的游程, 在这个序列中一共有 3 个"1"的游程. 我们感兴趣于计算游程的个数的期望值. 记 $R(1)$ 为排列中"1"的游程的个数, 记

$$I_i = \begin{cases} 1 & \text{一个 1 的游程开始于第 } i \text{ 个位置} \\ 0 & \text{其他} \end{cases}$$

这样 $R(1)$ 可表示成: $R(1) = \sum_{i=1}^{n+m} I_i$, 从而

$$E[R(1)] = \sum_{i=1}^{n+m} E[I_i]$$

经计算,

$$E[I_1] = P\{\text{"1" 在第一位置}\} = \frac{n}{n+m}$$

对于 $1 < i \leqslant n + m$,

$$E[I_i] = P\{\text{"0" 在第 } i-1 \text{ 个位置, "1" 在第 } i \text{ 个位置}\} = \frac{m}{n+m} \cdot \frac{n}{n+m-1}$$

这样,

$$E[R(1)] = \frac{n}{n+m} + (n+m-1)\frac{nm}{(n+m)(n+m-1)} = \frac{n(m+1)}{n+m}$$

类似地, 关于"0"的游程的个数的期望,

$$E[R(0)] = \frac{m(n+1)}{(n+m)}$$

以及关于总的游程的个数的期望,

$$E[R(1) + R(0)] = 1 + \frac{2mn}{m+n}$$ ■

例 2l (平面上的随机徘徊) 设在平面坐标系的原点上放一质点, 质点在平面上作如下的随机徘徊. 每一步质点移动一个单位距离, 且前进方向与 x 轴的夹角 θ 在 $(0, 2\pi)$ 上均匀分布. (见图 7.1.) 计算在走了 n 步以后, 质点离开原点的距离的平方的期望值.

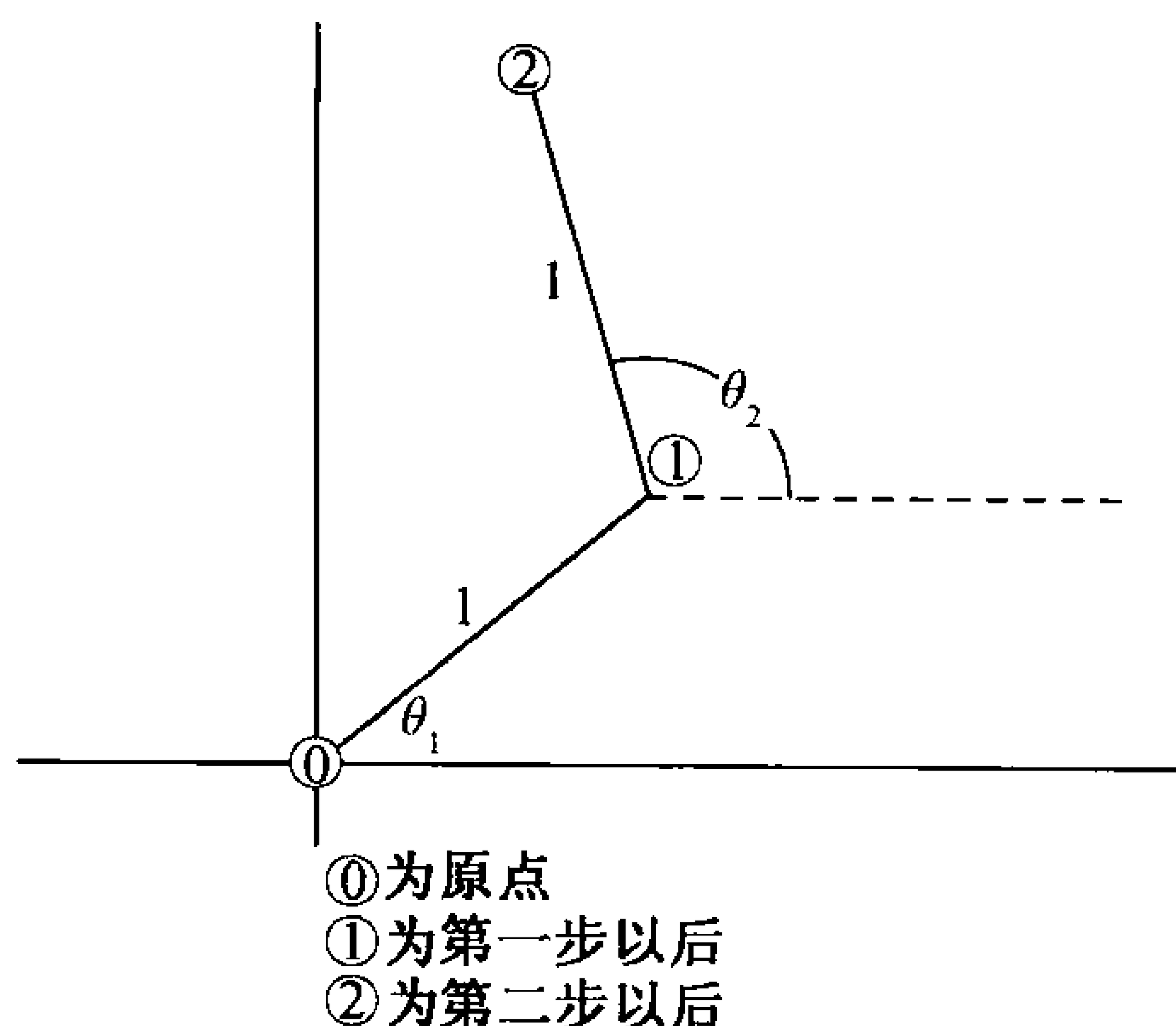


图7.1 平面上的随机徘徊

解: 用 (X_i, Y_i) 表示第 i 步移动后, 坐标的变动量, 因此

$$X_i = \cos\theta_i \qquad Y_i = \sin\theta_i$$

其中 $\theta_i, i = 1, 2, \cdots, n$ 为相互独立, 且在 $(0, 2\pi)$ 上均匀分布. 经过 n 步以后, 质点的位置为 $(\sum_{i=1}^n X_i, \sum_{i=1}^n Y_i)$, 质点离开原点的距离的平方 D^2 的公式为

$$\begin{aligned} D^2 &= \Big(\sum_{i=1}^n X_i\Big)^2 + \Big(\sum_{i=1}^n Y_i\Big)^2 = \sum_{i=1}^n (X_i^2 + Y_i^2) + \sum_{i \neq j}\sum (X_i X_j + Y_i Y_j) \\ &= n + \sum_{i \neq j}\sum (\cos\theta_i \cos\theta_j + \sin\theta_i \sin\theta_j) \end{aligned}$$

注意, 此处利用了公式 $\cos^2\theta_i + \sin^2\theta_i = 1$. 在求 D^2 的期望时, 利用 θ_i 之间相互独立性假设以及

$$2\pi E[\cos\theta_i] = \int_0^{2\pi} \cos u du = \sin 2\pi - \sin 0 = 0$$

$$2\pi E[\sin\theta_i] = \int_0^{2\pi} \sin u du = \cos 0 - \cos 2\pi = 0$$

最后得到 $E[D^2] = n$. ■

例 2m (快速排序算法) 假设有一组互不相同的数 $x_1, \cdots, x_n$. 我们需要将它们排成一个上升的序列, 称这一过程为排序. 通常采用一个快速排序程序来完成这一任务, 其方法如下: 当 $n = 2$ 时, 只需直接比较这两个数并且将他们排成升序即可. 当 $n > 2$ 时, 随机地选一个数, 设为 x_i, 然后将所有其余的数与 x_i 作比较, 将小于 x_i 的数归入 x_i 左边一集合, 将大于 x_i 的数归入 x_i 右边一集合, 然后对于那些数的集合重复刚才的处理过程, 直到所有的数都已排成升序为止. 例如, 下面是 10 个不同的数

$$5, 9, 3, 10, 11, 14, 8, 4, 17, 6$$

开始时从中随机地选一个数 (每一个数被选中的概率都为 1/10) 比如说选中了 10, 然后每一个数与 10 做比较, 这样得到

$$\{5, 9, 3, 8, 4, 6\}, 10, \{11, 14, 17\}$$

现在我们对于不是单点集的数集进行再分类. 例如先对上述左边的集合进行再分类, 随机地选定一个数, 例如 6 被选定, 然后将数集中其余的数与 6 作比较, 得到

$$\{5,3,4\},6,\{9,8\},10,\{11,14,17\}$$

现在, 我们还是考虑最左边的非单点集合, 将它的元素进行比较. 最左边的非单点集合为 $\{5,3,4\}$, 从中随机地取出一个数, 例如 4, 这样, 经过排列以后得到

$$\{3\},4,\{5\},6,\{9,8\},10,\{11,14,17\}$$

重复这个过程直到每个花括弧中只有一个数为止. 这样快速排序法也就完成了.

在排序过程中, 最基本的运算是比较两个数的大小. 记 X 为实现排序所需的比较的次数, 则 $E[X]$ 是一个排序算法的效率的一个度量, 为计算 $E[X]$, 我们首先将 X 写成一列随机变量的和. 为解决这个问题, 首先将最小值命名为 1, 将第二小的值命名为 2, ……, 将最大值命名为 n, 对于 $i<j$, 记

$$I(i,j)=\begin{cases}1 & \text{在排序过程中, } i,j \text{ 直接比较过}\\ 0 & \text{其他}\end{cases}$$

由排列过程可知, 任意两个数 i 和 j 在排序过程中, 有可能从未比较过, 但若它们直接比较过, 只可能比较一次, 不可能再次比较, 这样, 由 X 的定义可知:

$$X=\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}I(i,j)$$

由此推得

$$\begin{aligned}E[X]&=E\Big[\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}I(i,j)\Big]=\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}E[I(i,j)]\\&=\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}P\{i \text{ 和 } j \text{ 在排序过程中直接进行比较过}\}\end{aligned}$$

现在需要计算概率

$$P\{i \text{ 和 } j \text{ 在排序过程中直接比较过}\}=\frac{2}{j-i+1}$$

考虑数集合 $i,i+1,\cdots,j$, 一共有 $j-i+1$ 个元素. 最初它们全包含在一个大的集合中, 用一个花括弧将这个大集合括起来. 在排序过程中, 随机地选定一个比较点, 若这个比较点不在此区间内, 即比 i 小, 或比 j 大, 则此时数 i,j 不会直接进行比较, 在比较后重新排序时, 它们还是处于同一花括弧内. 现在假定选择的比较点落入集合 $i,i+1,\cdots,j-1,j$ 之中, 则只有当选择点为 i 或 j 时, i 和 j 才会直接进行比较, 否则它们与比较点比较后, 将分别放进左、右的花括弧内, 而不会直接相互比较. 因此, i,j 直接进行比较的概率, 就是从 $i,i+1,\cdots,j-1,j$ 中选取 i 或 j 的概

率, 等于 $2/(j-i+1)$.

将得到的公式代入 $E[X]$ 的表达式后, 得到 $E[X]=\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\dfrac{2}{j-i+1}$. 现在当 n 充分大时, 利用近似公式

$$\sum_{j=i+1}^{n}\frac{2}{j-i+1}\approx\int_{i+1}^{n}\frac{2}{x-i+1}\mathrm{d}x=2\ln(x-i+1)\Big|_{i+1}^{n}$$
$$=2\ln(n-i+1)-2\ln 2\approx 2\ln(n-i+1)$$

这样,

$$E[X]\approx\sum_{i=1}^{n-1}2\ln(n-i+1)\approx 2\int_{1}^{n-1}\ln(n-x+1)\mathrm{d}x=2\int_{2}^{n}\ln y\ \mathrm{d}y$$
$$=2(y\ln y-y)\Big|_{2}^{n}\approx 2n\ln n$$

由此可知, 利用快速排序法, 当 n 充分大时, 大概进行了 $2n\ln n$ 次比较, 可将 n 个数进行排序. ■

例 2n (事件和的概率) 记 $A_1,\cdots,A_n$ 为 n 个事件, $X_1,\cdots,X_n$ 是它们的示性函数

$$X_i=\begin{cases}1 & A_i\text{发生}\\ 0 & \text{其他}\end{cases}$$

注意

$$1-\prod_{i=1}^{n}(1-X_i)=\begin{cases}1 & \bigcup A_i\text{发生}\\ 0 & \text{其他}\end{cases}$$

因此

$$E\Big[1-\prod_{i=1}^{n}(1-X_i)\Big]=P\Big(\bigcup_{i=1}^{n}A_i\Big)$$

将上式左边展开

$$P\Big(\bigcup_{i=1}^{n}A_i\Big)=E\Big[\sum_{i=1}^{n}X_i-\sum_{i<j}\sum X_iX_j+\sum_{i<j<k}\sum\sum X_iX_jX_k$$
$$-\cdots+(-1)^{n+1}X_1\cdots X_n\Big] \tag{2.3}$$

由

$$X_{i_1}X_{i_2}\cdots X_{i_k}=\begin{cases}1 & A_{i_1}A_{i_2}\cdots A_{i_k}\text{发生}\\ 0 & \text{其他}\end{cases}$$

可知 $E[X_{i_1}\cdots X_{i_k}]=P(A_{i_1}\cdots A_{i_k})$, (2.3) 式变成大家熟知的事件和容斥公式

$$P\left(\bigcup A_i\right)=\sum_i P(A_i)-\sum\sum_{i<j}P(A_iA_j)+\sum\sum\sum_{i<j<k}P(A_iA_jA_k)$$
$$-\cdots+(-1)^{n+1}P(A_1\cdots A_n)$$ ■

现设 $X_i(i\geqslant 1)$ 是一个随机变量序列, 每一项具有有限期望, 下面的公式却不是无条件成立的,

$$E\left[\sum_{i=1}^{\infty}X_i\right]=\sum_{i=1}^{\infty}E[X_i] \tag{2.4}$$

在下列的一系列等式中, 打问号"?"的一处是 (2.4) 式成立的关键.

$$\begin{aligned}E\left[\sum_{i=1}^{\infty}X_i\right]&=E\left[\lim_{n\to\infty}\sum_{i=1}^{n}X_i\right]\\&\overset{?}{=}\lim_{n\to\infty}E\left[\sum_{i=1}^{n}X_i\right]=\lim_{n\to\infty}\sum_{i=1}^{n}E[X_i]=\sum_{i=1}^{\infty}E[X_i]\end{aligned} \tag{2.5}$$

问号是指期望运算和极限运算是否可以交换的问题. 若两个运算可交换, 则 (2.4) 式成立. 在一般情况下, 这种可交换性是没有得到证实的. 但是下面两种重要的特殊情况下, (2.5) 式中的两种运算的可交换性是得到保障的.

1. X_i 均为非负随机变量 (即 $P\{X_i\geqslant 0\}=1$ 对一切 i 成立).

2. $\sum_{i=1}^{\infty}E[|X_i|]<\infty$.

例 2o 考虑非负整数值随机变量 X, 对一切 $i\geqslant 1$, 定义

$$X_i=\begin{cases}1 & X\geqslant i\\0 & X<i\end{cases}$$

我们得到

$$\sum_{i=1}^{\infty}X_i=\sum_{i=1}^{X}X_i+\sum_{i=X+1}^{\infty}X_i=\sum_{i=1}^{X}1+\sum_{i=X+1}^{\infty}0=X$$

由于 X_i 都是非负随机变量, 因此

$$E[X]=\sum_{i=1}^{\infty}E(X_i)=\sum_{i=1}^{\infty}P\{X\geqslant i\} \tag{2.6}$$

这是一个非常有用的恒等式. ■

例 2p 设有 n 个对象, 记为 $1,2,\cdots,n$. 这 n 个对象被放在计算机的 n 个单元中, 他们顺次放在序号为 $i_1,\cdots,i_n$ 的单元中, 其中 $\{i_1,\cdots,i_n\}$ 为 $\{1,2,\cdots,n\}$ 的一个排列. 现设每次从这 n 个对象中随机访问一个对象, 并且每次访问与过去的访问历史是相互独立的. 现设访问对象 i 的概率为 $P(i),\sum_{i=1}^{n}P(i)=1$, 假定 $P(i)$ 为已知. 现在的问题是什么样的单元摆放次序, 能够使得每次访问的对象所处的单元序号的期望值达到最小.

解: 不妨设这 n 个对象的访问概率满足条件 $P(1) \geqslant P(2) \geqslant \cdots \geqslant P(n)$. 我们指出这些对象的最优摆放次序应为 $O_0 = \{1, 2, \cdots, n\}$. 记 X 为被访对象在计算机中所处的单元的序号, 对于对象 $1, 2, \cdots, n$ 在计算机中的单元任一排序 $O = \{i_1, \cdots, i_n\}$:

$$P_O\{X \geqslant k\} = \sum_{j=k}^{n} P(i_j) \geqslant \sum_{j=k}^{n} P(j) = P_{1,2,\cdots,n}\{X \geqslant k\}$$

由此利用公式 (2.6), 可得 $E_O(X) \geqslant E_{1,2,\cdots,n}(X)$. 本问题的结论指出应该把最经常访问的对象放在最容易访问的单元. 这样的系统设计, 操作起来最顺手, 访问的期望时间也最短. ■

*7.2.1 通过概率方法将期望值作为界

通过在集合上引入概率, 可分析集合中元素的性质. 在第 3 章例 4l 中已经看到了该方法的应用. 在本节中, 我们将找出某些函数的界.

设 f 为定义在有限集 $\mathcal{S}$ 上的函数, 我们对该函数的最大值感兴趣

$$m = \max_{s \in \mathcal{S}} f(s)$$

令 S 为取值于 $\mathcal{S}$ 的随机元, 显然, 由 $m \geqslant f(S)$ 知 $m \geqslant E[f(S)]$. 当 $f(S)$ 不是取常数值时, 上述不等号是严格意义下的不等号, 即不会出现 $m = E[f(S)]$ 的情况. 总之, $E[f(S)]$ 是 m 的下界.

例 2q (竞赛中的哈密顿路径数的最大值) 在循环赛中, n 个竞争对手中任意一对都进行比赛, 一共进行 $\binom{n}{2}$ 场比赛, 现将运动员编成号 $1, 2, \cdots, n$, 下面的情形形成哈密顿路径: $i_1, \cdots, i_n$, 如果 i_1 胜 i_2, i_2 胜 i_3, 如此下去, 直到 i_{n-1} 胜 i_n. 一个问题是求哈密顿路径的最大可能数.

作为解释, 设 $n = 3$, 如果有一个人胜了 2 次, 此时, 只可能有一条哈密顿路径, 例如, 1 胜了 2 和 3, 2 胜了 3, 则唯一的哈密顿路径为 1, 2, 3. 另外, 若在循环赛中没有人赢得 2 次, 即每人胜负各 1 次, 1 胜 2, 2 胜 3, 3 胜 1, 此时, 哈密顿路径的条数的最大值为 3.

我们将指出对于参赛人数为 n 的比赛中, 哈密顿路径可多于 $n!/2^{n-1}$ 条. 由于一共要进行 $\binom{n}{2}$ 场比赛, 比赛的结果一共有 $2^{\binom{n}{2}}$ 种, 记 $\mathcal{S}$ 为这 $2^{\binom{n}{2}}$ 个不同的比赛结果之集合, 记 $f(s)$ 表示比赛结果为 s 时的哈密顿路径的数目, $s \in \mathcal{S}$, 为证明 $\max\limits_{s} f(s) \geqslant \dfrac{n!}{2^{n-1}}$, 我们随机地抽取一比赛结果 s, $s \in \mathcal{S}$, 且 S 取每一个 $s \in \mathcal{S}$ 的概率都是相等的, 并假定这 $\binom{n}{2}$ 场比赛是相互独立的, 每个选手对等可能地打败对方.

现在将 n 个竞争者进行排队，一共有 $n!$ 种不同的排列，例如 $i_1, i_2, \cdots, i_n$ 表示竞争者的排列，其中 i_1 排第一位，i_2 排第二位 $\cdots$，将这些排列进行编号，$i = 1, 2, \cdots, n!$. 为了简化叙述，我们用 $i_1 \to i_2$ 表示 i_1 在与 i_2 的比赛中胜出. 对于确定的比赛结果，只有 $i_1 \to i_2 \to \cdots \to i_n$，排列 $(i_1, \cdots, i_n)$ 才是哈密顿路径. 记

$$X_i = \begin{cases} 1 & \text{若竞赛的结果使排列 } i \text{ 成为哈密顿路径} \\ 0 & \text{其他} \end{cases}$$

显然 X_i 依赖于比赛结果 s，是一随机变量，并且哈密顿路径数

$$f(S) = \sum_{i=1}^{n!} X_i$$

故

$$E[f(S)] = \sum_{i=1}^{n!} E[X_i]$$

对于 X_i，其中 i 为某一排列，不妨设 i 为 $(1, 2, \cdots, n)$，$(1, 2, \cdots, n)$ 成为哈密顿路径只有 1 胜 2，2 胜 3，$\cdots$，故 $E[X_i] = P\{X_i = 1\} = 1/2^{n-1}$. 由此可知 $E[f(S)] = n!(1/2)^{n-1}$. 由于 $f(S)$ 不等于常数，故存在一个试验结果其哈密顿路径数超过 $n!/2^{n-1}$. ■

例 2r　一共有 52 棵树排列在一个圆周上，有 15 只金花鼠，生活在这些树上，指出存在相连的 7 棵树，其上生活着至少有 3 只金花鼠.

解：将一棵树和顺时针方向的另外 6 棵相邻的树组成该树的一个邻域. 现在我们需要证明：对于生活在这 52 棵树上的 15 只金花鼠，不论它们如何分布，我们总能找到一棵树，使得至少有 3 只金花鼠生活在这个树和它的邻域中. 为此，我们随机地选一棵树，并且数一数这棵树和它的邻域中的金花鼠的数目，记这个数为 X. 显然，X 是一个随机变量，现在将这 15 只金花鼠记为 $i = 1, 2, \cdots, 15$，记

$$X_i = \begin{cases} 1 & \text{金花鼠 } i \text{ 在这个随机选定的一棵树的邻域里} \\ 0 & \text{其他} \end{cases}$$

显然 $X = \sum_{i=1}^{15} X_i$，从而 $E[X] = \sum_{i=1}^{15} E[X_i]$. $\{X_i = 1\}$ 表示随机地选定的 7 棵树上生活着金花鼠 i，显然 $E[X_i] = P\{X_i = 1\} = 7/52$，从而 $E[X] = 15 \times (7/52) = 105/52 > 2$. 由此可知，必定存在一棵树，使得在这棵树的邻域里生活的金花鼠数目大于 2. ■

*7.2.2　关于最大数与最小数的恒等式

下面是关于若干数的最大值，最小值的一个恒等式.

命题 2.2 对于任意一组数 $x_1,\cdots,x_n$, 下列恒等式成立:

$$\max_i x_i = \sum_i x_i - \sum_{i<j}\min(x_i,x_j) + \sum_{i<j<k}\min(x_i,x_j,x_k) + \cdots + (-1)^{n+1}\min(x_1,\cdots,x_n)$$

证明: 我们给出一个概率的证明. 首先, 假定所有的 x_i 在 $[0,1]$ 区间内, 令 U 为 $(0,1)$ 上均匀分布的随机变量, 记事件 $A_i=\{U<x_i\}$, 不难验证

$$\bigcup_i A_i = \{U<\max_i x_i\}$$

因此,

$$P\{\bigcup_i A_i\} = P\{U<\max_i x_i\} = \max_i x_i$$

及

$$P(A_i) = P\{U<x_i\} = x_i$$

此外, 事件 $A_{i_1},\cdots,A_{i_r}$ 同时发生的充要条件是事件 $\{U<x_{i_j}\}, j=1,\cdots,r$ 同时成立, 由此导出

$$A_{i_1}\cdots A_{i_r} = \{U<\min_{j=1,\cdots,r} x_{i_j}\}$$

进一步得

$$P(A_{i_1}\cdots A_{i_r}) = P\{U<\min_{j=1,\cdots,r} x_{i_j}\} = \min_{j=1,\cdots,r} x_{i_j}$$

由事件和的概率的容斥公式:

$$P\Big(\bigcup_i A_i\Big) = \sum_i P(A_i) - \sum_{i<j}P(A_iA_j) + \sum_{i<j<k}P(A_iA_jA_k) + \cdots + (-1)^{n+1}P(A_1\cdots A_n)$$

可得到该恒等式.

现在设 x_i 为非负, 但不限制于单位区间. 设 c 为常数, 使所有 x_i 均小于 c, 此时恒等式对于 $y_i=x_i/c$ 成立, 再在恒等式两边乘以常数 c, 可知恒等式对 x_i 也成立. 现在假定 x_i 可取负值, 此时, 存在 b, 使得 $x_i+b>0$ 对一切 $i=1,\cdots,n$ 成立. 这样, 下列恒等式成立

$$\max_i(x_i+b) = \sum_i(x_i+b) - \sum_{i<j}\min(x_i+b,x_j+b) + \cdots + (-1)^{n+1}\min(x_1+b,\cdots,x_n+b)$$

记 $M=\sum_i x_i - \sum_{i<j}\min(x_i,x_j)+\cdots+(-1)^{n+1}\min(x_1,\cdots,x_n)$, 可将上面的恒等式写成:

$$\max_i x_i + b = M + b\Big(n - \binom{n}{2} + \cdots + (-1)^{n+1}\binom{n}{n}\Big)$$

但是下面也是一个恒等式

$$0=(1-1)^n=1-n+\binom{n}{2}+\cdots+(-1)^n\binom{n}{n}$$

将上面的两个恒等式合并便得

$$\max_i x_i = M$$

这样, 证明了本命题所列的恒等式对一切 $x_i(i=1,\cdots,n)$ 都成立. □

由 命题 2.2 知, 对任意随机变量 $X_1,\cdots,X_n$, 有

$$\max_i X_i=\sum_i X_i-\sum_{i<j}\min(X_i,X_j)+\cdots+(-1)^{n+1}\min(X_1,\cdots,X_n)$$

将上式求期望, 可得

$$E[\max_i X_i]=\sum_i E[X_i]-\sum_{i<j}E[\min(X_i,X_j)] \\ +\cdots+(-1)^{n+1}E[\min(X_1,\cdots,X_n)] \quad (2.7)$$

例 2s (不等概率的优惠券收集问题) 现设某人收集优惠券, 每次收集一张, 均独立于以前所收集的优惠券. 现设一共有 n 种不同的优惠券, 每次收集的时候, 优惠券 i 被收集到的概率为 $p_i, \sum_{i=1}^n p_i=1$. 求出当第一次收集到全套 n 种优惠券时所收集到优惠券总数的期望值.

解: 记 X_i 为收集到第 i 种优惠券时所收集到的优惠券张数, 记 X 为收集到全套 n 种优惠券时所收集到的优惠券总数. X 与 X_i 之间有如下的关系

$$X=\max_{i=1,\cdots,n} X_i$$

由于收集到一张第 i 种优惠券的概率为 p_i, X_i 的分布是以 p_i 为参数的几何分布. 又由于 $\min\{X_i,X_j\}$ 是为了收集到第 i 种优惠券或第 j 种优惠券所需收集的优惠券的张数, 因此 $\min(X_i,X_j)$ 的分布是以 p_i+p_j 为参数的几何分布. 类似地, $\min(X_i,X_j,X_k)$ 的分布是以 $p_i+p_j+p_k$ 为参数的几何分布, 利用 (2.7) 式以及几何分布的性质可知

$$E[X]=\sum_i\frac{1}{p_i}-\sum_{i<j}\frac{1}{p_i+p_j}+\sum_{i<j<k}\frac{1}{p_i+p_j+p_k} \\ +\cdots+(-1)^{n+1}\frac{1}{p_1+\cdots+p_n}$$

注意到

$$\int_0^\infty e^{-px}dx=\frac{1}{p}$$

并利用恒等式

$$1-\prod_{i=1}^n(1-e^{-p_ix})=\sum_i e^{-p_ix}-\sum_{i<j}e^{-(p_i+p_j)x}+\cdots+(-1)^{n+1}e^{-(p_1+\cdots+p_n)x}$$

关于 $E[X]$ 的等式可以化成下面更便于计算的公式

$$E[X] = \int_0^\infty \left(1 - \prod_{i=1}^n (1 - \mathrm{e}^{-p_i x})\right) \mathrm{d}x$$ ■

7.3 试验序列中事件发生次数的矩

上一节中的许多例子都具有下列形式：对于给定的事件序列 $A_1, \cdots, A_n$，求出 $E[X]$，其中 X 是这些事件在试验中的发生次数. 其解法是给出每个事件的示性函数

$$I_i = \begin{cases} 1 & A_i\text{发生} \\ 0 & \text{其他} \end{cases}$$

利用关系式 $X = \sum_{i=1}^n I_i$，可得

$$E[X] = E\Big[\sum_{i=1}^n I_i\Big] = \sum_{i=1}^n E[I_i] = \sum_{i=1}^n P(A_i) \tag{3.1}$$

现在我们感兴趣于“事件对”出现的次数. 若 A_i 与 A_j 在试验中出现，则 $I_i \cdot I_j = 1$，反之，则 $I_i \cdot I_j = 0$，因此，在试验序列中，$\sum_{i<j} I_i I_j$ 是事件对出现的次数. 又由于 X 是试验序列中事件出现的次数，因此事件对出现的次数为 $\binom{X}{2}$. 这样，

$$\binom{X}{2} = \sum_{i<j} I_i I_j$$

在上式右边的和项中，一共有 $\binom{n}{2}$ 项. 上式两边求期望，得

$$E\Big[\binom{X}{2}\Big] = \sum_{i<j} E[I_i I_j] = \sum_{i<j} P(A_i A_j) \tag{3.2}$$

或

$$E\Big[\frac{X(X-1)}{2}\Big] = \sum_{i<j} P(A_i A_j)$$

由此得到

$$E[X^2] - E[X] = 2\sum_{i<j} P(A_i A_j) \tag{3.3}$$

进一步可得到 $E[X^2]$ 和 $\mathrm{Var}(X) = E[X^2] - (E[X])^2$ 的公式.

进一步，考虑在试验序列中，k 个事件组的出现次数，可得到

$$\binom{X}{k} = \sum_{i_1 < i_2 < \cdots < i_k} I_{i_1} I_{i_2} \cdots I_{i_k}$$

两边求期望得到

$$E\left[\binom{X}{k}\right]=\sum_{i_1<i_2<\cdots<i_k}E[I_{i_1}I_{i_2}\cdots I_{i_k}]=\sum_{i_1<i_2<\cdots<i_k}P(A_{i_1}A_{i_2}\cdots A_{i_k}) \tag{3.4}$$

例 3a (二项随机变量的矩) 考虑 n 次独立重复试验, 每次成功的概率为 p, 记 A_i 为第 i 次试验成功这一事件, 当 $i\neq j$ 时, $P(A_iA_j)=p^2$, 由 (3.2) 式得到

$$E\left[\binom{X}{2}\right]=\sum_{i<j}p^2=\binom{n}{2}p^2$$

或

$$E[X(X-1)]=n(n-1)p^2$$

或

$$E[X^2]-E[X]=n(n-1)p^2$$

再利用 $E[X]=\sum_{i=1}^{n}P(A_i)=np$, 可得

$$\text{Var}(X)=E[X^2]-(E[X])^2=n(n-1)p^2+np-(np)^2=np(1-p)$$

这个结果与 4.6.1 节的结果相同.

一般情况下, 利用 $P(A_{i_1}A_{i_2}\cdots A_{i_k})=p^k$, (3.4) 式变成

$$E\left[\binom{X}{k}\right]=\sum_{i_1<i_2<\cdots<i_k}p^k=\binom{n}{k}p^k$$

或等价地

$$E[X(X-1)\cdots(X-k+1)]=n(n-1)\cdots(n-k+1)p^k$$

利用上式可以递推得到各阶矩 $E[X^k], k\geqslant 3$. 例如, $k=3$ 的情况下,

$$E[X(X-1)(X-2)]=n(n-1)(n-2)p^3$$

或

$$E[X^3-3X^2+2X]=n(n-1)(n-2)p^3$$

从而

$$\begin{aligned}E[X^3]&=3E[X^2]-2E[X]+n(n-1)(n-2)p^3\\&=3n(n-1)p^2+np+n(n-1)(n-2)p^3\end{aligned}$$

■

例 3b (超几何随机变量的矩) 设一个坛子中有 N 个球, 其中 m 个为白球, $N-m$ 个黑球. 现从中随机地抽取 n 个球, 此时 n 个球中的白球个数 X 就是事件 $A_1, A_2, \cdots, A_n$ 的发生数, 其中事件 A_i 表示取出的第 i 个球为白球. 由于第 i 个球可以是 N 个球中的任意一个, 其中 m 个为白球, 因此 $P(A_i)=m/N$. 由公式 (3.1) 得到 $E[X]=\sum_{i=1}^{n}P(A_i)=nm/N$. 又由于

$$P(A_iA_j)=P(A_i)P(A_j|A_i)=\frac{m}{N}\cdot\frac{m-1}{N-1}$$

再利用 (3.2) 式得

$$E\left[\binom{X}{2}\right]=\sum_{i<j}\frac{m(m-1)}{N(N-1)}=\binom{n}{2}\frac{m(m-1)}{N(N-1)}$$

或

$$E[X(X-1)]=n(n-1)\cdot\frac{m(m-1)}{N(N-1)}$$

或

$$E[X^2]=n(n-1)\frac{m(m-1)}{N(N-1)}+E[X]$$

进一步可得

$$\begin{aligned}\text{Var}(X)=E[X^2]-(E[X])^2&=n(n-1)\frac{m(m-1)}{N(N-1)}+\frac{nm}{N}-\frac{n^2m^2}{N^2}\\&=\frac{mn}{N}\left[\frac{(n-1)(m-1)}{N-1}+1-\frac{mn}{N}\right]\end{aligned}$$

这个结果与第 4 章例 8j 所得的结果是相同的.

利用 (3.4) 式, 可得到 X 的高阶矩. 由

$$P(A_{i_1}A_{i_2}\cdots A_{i_k})=\frac{m(m-1)\cdots(m-k+1)}{N(N-1)\cdots(N-k+1)}$$

再利用 (3.4) 式, 得

$$E\left[\binom{X}{k}\right]=\binom{n}{k}\cdot\frac{m(m-1)\cdots(m-k+1)}{N(N-1)\cdots(N-k+1)}$$

或

$$E[X(X-1)\cdots(X-k+1)]=n(n-1)\cdots(n-k+1)\cdot\frac{m(m-1)\cdots(m-k+1)}{N(N-1)\cdots(N-k+1)}$$ ■

例 3c (配对问题中的矩) 设有 N 个人, 随机地拿帽子. 记 A_i 为第 i 个人拿到自己的帽子. 此时

$$P(A_iA_j)=P(A_i)P(A_j|A_i)=\frac{1}{N}\cdot\frac{1}{N-1}$$

其中, $P(A_j|A_i)$ 表示第 i 个人已经拿到自己的帽子的条件下, 第 j 个人拿到自己帽子的概率. 这个概率与第 i 个人拿到自己帽子的概率相似, 只是可选择帽子总数为 $N-1$. 记 X 为 n 个人中拿到自己帽子的人数, 由公式 (3.2) 得到

$$E\left[\binom{X}{2}\right]=\sum_{i<j}\frac{1}{N(N-1)}=\binom{N}{2}\frac{1}{N(N-1)}$$

从而, $E[X(X-1)]=1$, 也即, $E[X^2]=1+E[X]$. 由于 $E[X]=\sum_{i=1}^{N}P(A_i)=1$, 我们得到

$$\text{Var}(X)=E[X^2]-(E[X])^2=1$$

这样, X 的期望和方差均为 1. 对于高阶矩, 可利用 (3.4) 式, 由于 $P(A_{i_1}A_{i_2}\cdots A_{i_k})=1/[N(N-1)\cdots(N-k+1)]$, 可得

$$E\left[\binom{X}{k}\right] = \binom{N}{k} \cdot \frac{1}{N(N-1)\cdots(N-k+1)}$$

或

$$E[X(X-1)\cdots(X-k+1)] = 1$$ ■

例 3d (另一个优惠券收集问题)　设有 N 种不同的优惠券, 每次收集到的新优惠券均与以前收集到的优惠券相互独立. 假定得到优惠券 i 的概率为 p_i, $\sum_{i=1}^{N} p_i = 1$. 找出当收集到 n 张优惠券时, 不同类型优惠券的类别数的期望与方差.

解: 我们发现讨论未收集到的优惠券类别更方便. 令 Y 为已收集到的类别数, $X = N - Y$ 为未收集到的类别数, 记 A_i 表示收集到的 n 张优惠券中没有类别 i 这一事件, 因此, X 是 $A_1, \cdots, A_N$ 中发生的事件数, 每次收集到 i 以外的优惠券的概率为 $1 - p_i$. 因此, $P(A_i) = (1-p_i)^n$. 从而 $E[X] = \sum_{i=1}^{N}(1-p_i)^n$, 我们得到

$$E[Y] = N - E[X] = N - \sum_{i=1}^{N}(1-p_i)^n$$

类似地, 在收集优惠券时, 既没有 i, 又没有 j 的概率为 $1 - p_i - p_j$. 同时, 每次收集的优惠券种类与以前所收集到的优惠券相互独立, 因此

$$P(A_iA_j) = (1-p_i-p_j)^n, \qquad i \neq j$$

这样,

$$E[X(X-1)] = 2\sum_{i<j} P(A_iA_j) = 2\sum_{i<j}(1-p_i-p_j)^n$$

或

$$E[X^2] = 2\sum_{i<j}(1-p_i-p_j)^n + E[X]$$

这样, 我们得到

$$\begin{aligned}
\mathrm{Var}(Y) = \mathrm{Var}(X) &= E[X^2] - (E[X])^2 \\
&= 2\sum_{i<j}(1-p_i-p_j)^n + \sum_{i=1}^{N}(1-p_i)^n - \left(\sum_{i=1}^{N}(1-p_i)^n\right)^2
\end{aligned}$$

特别, 当 $p_i = 1/N$, $i = 1, 2, \cdots, N$ 时,

$$E[Y] = N\left[1 - \left(1 - \frac{1}{N}\right)^n\right]$$

$$\mathrm{Var}(Y) = N(N-1)\left(1 - \frac{2}{N}\right)^n + N\left(1 - \frac{1}{N}\right)^n - N^2\left(1 - \frac{1}{N}\right)^{2n}$$ ■

例 3e (负超几何分布随机变量)　**负超几何分布**(negative hypergeometric distribution) 是这样定义的: 设一坛子内共有 $m + n$ 个球, 其中 n 个为特别的球, m 个为通常的球, 每次从坛子中随机地取出一个球, 即坛子中的每一个球都以相等的概率被取到. 记 Y 为取到 r 个特别球的时候所需的抽取次数, 则 Y 的分布就是负

超几何分布. 负超几何分布与超几何分布的关系同负二项分布与二项分布的关系是一样的. 它们不是考虑成功次数的概率, 而是考虑为达到固定成功次数所需试验次数的分布. 现在求 Y 的分布. $Y=k$ 是这样的事件, 它是由下列两个条件所确定的:

(a) 前面 $k-1$ 个球中, 含 $r-1$ 个特殊的球, $k-r$ 个通常的球.

(b) 第 k 个球为特殊球.

显然,

$$P\{Y=k\}=\frac{\binom{n}{r-1}\binom{m}{k-r}}{\binom{n+m}{k-1}}\cdot\frac{n-r+1}{n+m-k+1}$$

我们不打算用 Y 的分布直接计算 Y 的方差和期望. 我们采用如下的技巧: 将 m 个通常的球记成 $o_1,\cdots,o_m$, 用 A_i 表示"在 r 个特殊的球被取走以前, o_i 已经被取走"这一事件, 令 X 为 $A_1,\cdots,A_m$ 中事件的发生数, 则 X 就是 r 个特殊的球被取走时被取走的通常球的个数. 显然 $Y=r+X$, 因此,

$$E[Y]=r+E[X]=r+\sum_{i=1}^{m}P(A_i)$$

为计算 $P(A_i)$, 考虑 $n+1$ 个球, 其中一个球为 o_i, 其他的球都是特殊的球. 设想从坛子里一个一个地往外随机地取球, 最后把全部球都取出来. 现在可产生 $n+1$ 种情况, 一种情况是 o_i 比这 n 个特殊球早出来, 第 2 种情况是最先取到一个特殊球, 然后出现 o_i, 然后其他球陆续出来, 等等, 最后一种情况是所有 n 个特殊球都取出以后, o_i 才被取出来.

由于这 $n+1$ 个球都处于平等的地位, 上述的 $n+1$ 种情况是等可能的, 因此, 每种情况出现的概率为 $1/n$, 而 A_i 刚好是由前 r 种情况所组成, 因此

$$P(A_i)=\frac{r}{n+1}$$

这样

$$E[Y]=r+m\cdot\frac{r}{n+1}=\frac{r(n+m+1)}{n+1}$$

举例来说, 在翻牌游戏中, 在一副扑克牌中, 翻出一张黑桃所需的平均翻牌数为 $1+39/14=3.786(r=1,n=13,m=39)$, 而翻出一张"A"所需的平均张数为 $1+48/5=10.6(r=1,n=4,m=48)$.

现在计算 $\mathrm{Var}(Y)=\mathrm{Var}(X)$. 我们利用等式

$$E[X(X-1)]=2\sum_{i<j}P(A_iA_j)$$

其中 $P(A_iA_j)$ 是 r 个特殊球被取出以前 o_i,o_j 已经被取出来的概率. 考虑 $n+2$ 个球, 其中包括 o_i,o_j 以及 n 个特殊的球. 由于 o_i,o_j 以及 n 个特殊的球在抽取过程

中完全处于平等地位, 因此事件 A_iA_j 等价于 o_i 和 o_j 这两个球在被取出的过程中排在前 $r+1$ 位, 这样,

$$P(A_iA_j) = \binom{2}{2}\binom{n}{r-1}\Big/\binom{n+2}{r+1} = \frac{r(r+1)}{(n+1)(n+2)}$$

从而,

$$E[X(X-1)] = 2\binom{m}{2}\frac{r(r+1)}{(n+1)(n+2)}$$

或

$$E[X^2] = m(m-1)\frac{r(r+1)}{(n+1)(n+2)} + E[X]$$

由于 $E[X] = mr/(n+1)$, 我们有

$$\begin{aligned}\text{Var}(Y) = \text{Var}(X) &= m(m-1)\frac{r(r+1)}{(n+1)(n+2)} \cdot \frac{mr}{n+1} - \left(\frac{mr}{n+1}\right)^2 \\ &= \frac{mr(n+1-r)(n+m+1)}{(n+1)^2(n+2)}\end{aligned}$$ ■

例 3f (优惠券收集中落单种类问题) 设一共有 n 种不同的优惠券, 在收集优惠券时, 假设各种优惠券是等可能出现的, 并且与以前收集的历史是相互独立的. 现设一个人不断收集优惠券, 直到全部 n 种优惠券收集齐全为止. 现在希望求出在收集过程中, 落单的优惠券的种类数的期望和方差.①

解: 记 X 表示在收集过程中落单的优惠券数目. 令 T_i 表示收集到的第 i 种优惠券的种类.② 用 A_i 表示 T_i 在收集的过程中落单这一事件, 则 X 等于 $A_1 \cdots, A_n$ 中发生的事件数, 从而 $E[X] = \sum_{i=1}^{n} P(A_i)$.

现在求 $P(A_i)$. 当第一次收集到 T_i 类型的优惠券后还需收集 $(n-i)$ 种新的类型的优惠券.③ 此时, 这 $n-i+1$ 种 ($n-i$ 种未收集的以及 T_i) 优惠券同等可能

---

① 一种优惠券称为落单的, 是指当收集到全部种类的优惠券时, 这种优惠券只收集到一张. —— 译者注

② 例如, 第一次收集到的优惠券的种类称为 T_1, 因为它是收集到的第 1 种优惠券. 若第二次收集到的不是 T_1 类型, 则称为类型 T_2, 因为它是第 2 种优惠券. 因此, 收集到的优惠券序列总是 T_1, $T_1T_2T_3T_1 \cdots$.—— 译者注

③ 现在看第一次收集到 T_i 以后的优惠券收集情况. 我们只关心这 $n-i+1$ 种优惠券出现的情况, 其中 $n-i$ 种是以前没有收集到的优惠券, 还有一种就是 T_i. 我们将这 $n-i+1$ 种优惠券在收集中出现的顺序分成两类. 一种情况是 T_i 在最后出现, 这种序列是这样的: $T_i, \cdots W_{i_1} \cdots W_{i_2}, \cdots W_{j_{n-i}} \cdots T_i$, 这个序列的第一个总是 T_i, 表示在整个收集过程中第一次收集到第 i 种优惠券 T_i, 后面的 W_{j_1} 表示出现了一个新的优惠券种类. 此后出现的种类依次为 $W_{j_2}, \cdots$. 不过除了第一个为 T_i 外, 在 $W_{j_{n-i}}$ 以前没有第二个 T_i. 这个序列实际上到 $W_{j_{n-i}}$ 就终止了, 因为已经收集到全套优惠券了, 因此, 出现这种情况 T_i 就落单了. 另外一种情况为 T_i, $W_{j_1}, \cdots W_{j_{n-i+1}} W_{j_{n-i}} \cdots$ 而诸 W 之前除了第一个 T_i 以外, 还有一个 T_i, 此时, T_i 就不落单. (此处 $j_1, \cdots j_{n-i}$ 为 $1, \cdots n-i$ 的一个排列.) 这样看 T_i 是不是落单, 就看在第一次出现 T_i 以后, 这 $n-i+1$ 种优惠券哪一种最后出现, 若 T_i 最后出现, T_i 就落单. 若这 $(n-i)$ 种新的优惠券中的某一种最后出现, T_i 就不落单. —— 译者注

地能成为最后一类被收集到的优惠券. 显然 T_i 落单的概率为 $1/(n-i+1)$, 因此 $P(A_i)=1/(n-i+1)$, 从而

$$E[X]=\sum_{i=1}^{n}\frac{1}{n-i+1}=\sum_{i=1}^{n}\frac{1}{i}$$

现在确定落单优惠券数的方差. 对 $i<j$, 定义 S_{ij} 表示“第一次收集到 T_j 时, T_i 仍只有一张”, 显然 $A_iA_j\subseteq S_{ij}$,

$$P(A_iA_j)=P(A_iA_j|S_{ij})P(S_{ij})$$

$P(S_{ij})$ 是如下事件的概率: 收集到 T_i 后, 在 T_i 和待收集的 $n-i$ 种共 $n+1-i$ 种优惠券中, 收集的前 $j-i$ 种中没有 T_i. 由于 T_i 等可能地成为第一、第二、……、第 $n+1-i$ 种再次被收集到的品种, 因此

$$P(S_{ij})=1-\frac{j-i}{n-i+1}=\frac{n+1-j}{n+1-i}$$

现在考虑 S_{ij} 发生的条件下, A_iA_j 发生的概率. 当第一次收集到 T_j 以后 (T_i 在此前已经收集到), 考虑以后的 $n-j+2$ 种优惠券, 其中包括尚未收集到的 $n-j$ 种新的优惠券以及 T_i 和 T_j, 只有 T_i,T_j 出现在 $n-j$ 种未收集到的优惠券之后, 事件 A_iA_j 才发生. 由于这 $n-j+2$ 种优惠券的地位是平等的, 这样

$$P(A_iA_j|S_{ij})=1\Big/\binom{n+2-j}{2}=\frac{2}{(n-j+2)(n-j+1)}$$

由此可得

$$P(A_iA_j)=\frac{2}{(n-j+2)(n-i+1)}\qquad i<j$$

因此

$$E[X(X-1)]=4\sum_{i<j}\frac{1}{(n+1-i)(n+2-j)}$$

利用前面得到的 $E[X]$ 的公式, 有

$$\operatorname{Var}(X)=4\sum_{i<j}\frac{1}{(n+1-i)(n+2-j)}+\sum_{i=1}^{n}\frac{1}{i}-\Big(\sum_{i=1}^{n}\frac{1}{i}\Big)^2$$ ■

7.4 协方差、和的方差及相关系数

下面是关于独立随机变量乘积的期望的一个命题.

命题 4.1　若 X, Y 相互独立, 则对任何函数 h 和 g, 下式成立:

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$$

证明: 设 X, Y 具有联合密度 $f(x, y)$, 则

$$\begin{aligned}
E[g(X)h(Y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f(x,y)\mathrm{d}x\mathrm{d}y \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f_X(x)f_Y(y)\mathrm{d}x\mathrm{d}y \\
&= \int_{-\infty}^{\infty} h(y)f_Y(y)\mathrm{d}y \int_{-\infty}^{\infty} g(x)f_X(x)\mathrm{d}x \\
&= E[h(Y)]E[g(X)]
\end{aligned}$$

离散情形下的证明是完全类似的. □

期望和方差可以给出单个随机变量的信息, 两个随机变量的协方差可指示两个随机变量之间的关系.

定义　X 和 Y 之间的协方差 $\mathrm{Cov}(X, Y)$ 由下式给出:

$$\mathrm{Cov}(X,Y) = E[(X - E[X])(Y - E[Y])]$$

将上式右边期望号下的表达式展开, 可得

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= E[XY - E[X]Y - XE[Y] + E[X]E[Y]] \\
&= E[XY] - E[X]E[Y] - E[X]E[Y] + E[X]E[Y] \\
&= E[XY] - E[X]E[Y]
\end{aligned}$$

若 X, Y 相互独立, 则由命题 4.1 可知 $\mathrm{Cov}(X, Y) = 0$. 但是其逆命题却不真. 下面给出一个相依随机变量具有零协方差的例子. 设 X 满足下式:

$$P\{X = 0\} = P\{X = 1\} = P\{X = -1\} = 1/3$$

定义

$$Y = \begin{cases} 0 & X \neq 0 \\ 1 & X = 0 \end{cases}$$

由 $XY \equiv 0$ 可知 $E[XY] = 0$, 同时 $E[X] = 0$, 从而

$$\mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y] = 0$$

然而, X, Y 并不独立.

下面的命题列出了协方差的若干性质.

命题 4.2 (i) $\text{Cov}(X,Y)=\text{Cov}(Y,X)$ (ii) $\text{Cov}(X,X)=\text{Var}(X)$
(iii) $\text{Cov}(aX,Y)=a\text{Cov}(X,Y)$
(iv) $\text{Cov}\Big(\sum_{i=1}^n X_i,\sum_{j=1}^m Y_j\Big)=\sum_{i=1}^n\sum_{j=1}^m \text{Cov}(X_i,Y_j)$

命题 4.2 的证明: (i)(ii) 可直接由协方差的定义证得. (iii) 作为练习留给读者去证. 现证结论 (iv). 记 $\mu_i=E[X_i], v_j=E[Y_j]$. 则由

$$E\Big[\sum_{i=1}^n X_i\Big]=\sum_{i=1}^n\mu_i,\qquad E\Big[\sum_{j=1}^m Y_i\Big]=\sum_{j=1}^m\nu_j$$

得

$$\begin{aligned}\text{Cov}\Big(\sum_{i=1}^n X_i,\sum_{j=1}^m Y_j\Big)&=E\Big[\Big(\sum_{i=1}^n X_i-\sum_{i=1}^n\mu_i\Big)\Big(\sum_{j=1}^m Y_j-\sum_{j=1}^m\nu_j\Big)\Big]\\&=E\Big[\sum_{i=1}^n(X_i-\mu_i)\sum_{j=1}^m(Y_j-\nu_j)\Big]\\&=E\Big[\sum_{i=1}^n\sum_{j=1}^m(X_i-\mu_i)(Y_j-\nu_j)\Big]=\sum_{i=1}^n\sum_{j=1}^m E[(X_i-\mu_i)(Y_j-\nu_j)]\end{aligned}$$

上述最后一个等式也是利用了期望运算的可加性. □

利用命题 4.2的 (ii) 和 (iv), 并且取 $Y_j=X_j, j=1,\cdots,n$, 可得

$$\begin{aligned}\text{Var}\Big(\sum_{i=1}^n X_i\Big)&=\text{Cov}\Big(\sum_{i=1}^n X_i,\sum_{j=1}^n X_j\Big)=\sum_{i=1}^n\sum_{j=1}^n\text{Cov}(X_i,X_j)\\&=\sum_{i=1}^n\text{Var}(X_i)+\sum\sum_{i\neq j}\text{Cov}(X_i,X_j)\end{aligned}$$

在上式中, 对于 $i,j,i\neq j$, 二重和中出现了两次, 因此上式等价于

$$\text{Var}\left(\sum_{i=1}^n X_i\right)=\sum_{i=1}^n\text{Var}(X_i)+2\sum\sum_{i<j}\text{Cov}(X_i,X_j) \tag{4.1}$$

如果 $X_1,\cdots,X_n$ 两两独立, 即对于 $i\neq j$, X_i 与 X_j 相互独立. 此时, 方程 (4.1) 简化为:

$$\text{Var}\Big(\sum_{i=1}^n X_i\Big)=\sum_{i=1}^n\text{Var}(X_i)$$

下面的例子说明了公式 (4.1) 的用处.

例 4a 设 $X_1,\cdots,X_n$ 为独立同分布随机变量序列, 其公共期望为 μ, 方差为 σ^2. 如例 2c 那样, 令 $\bar{X}=\sum_{i=1}^n X_i/n$ 为样本均值, $X_i-\bar{X}, i=1,\cdots,n$ 称为离差, 等于个体数据与样本均值之差. 随机变量

$$S^2=\sum_{i=1}^n\frac{(X_i-\bar{X})^2}{n-1}$$

称为样本方差. 求 (a)$\text{Var}(\bar{X})$ 和 (b)$E[S^2]$.

解: (a) 利用独立性得

$$\text{Var}(\bar{X}) = \left(\frac{1}{n}\right)^2 \text{Var}\left(\sum_{i=1}^{n} X_i\right) = \left(\frac{1}{n}\right)^2 \sum_{i=1}^{n} \text{Var}(X_i) = \frac{\sigma^2}{n}$$

(b) 我们由下列代数恒等式开始计算

$$\begin{aligned}(n-1)S^2 &= \sum_{i=1}^{n}(X_i-\mu+\mu-\bar{X})^2 \\ &= \sum_{i=1}^{n}(X_i-\mu)^2 + \sum_{i=1}^{n}(\bar{X}-\mu)^2 - 2(\bar{X}-\mu)\sum_{i=1}^{n}(X_i-\mu) \\ &= \sum_{i=1}^{n}(X_i-\mu)^2 + n(\bar{X}-\mu)^2 - 2(\bar{X}-\mu)n(\bar{X}-\mu) \\ &= \sum_{i=1}^{n}(X_i-\mu)^2 - n(\bar{X}-\mu)^2\end{aligned}$$

再求期望, 得到

$$\begin{aligned}(n-1)E[S^2] &= \sum_{i=1}^{n} E[(X_i-\mu)^2] - nE[(\bar{X}-\mu)^2] \\ &= n\sigma^2 - n\text{Var}(\bar{X}) = (n-1)\sigma^2\end{aligned}$$

上面的最后等式是利用了 (a) 的结果, 而最前面的等式是利用例 2c(即 $\mu = E[\bar{X}]$) 的结果. 两边再除以 $(n-1)$ 就得到样本方差的期望为 σ^2. ■

下面例子提供了求二项随机变量方差的另一个方法.

例 4b (二项随机变量的方差) 设 X 为二项随机变量, 其参数为 (n,p), 求 $\text{Var}(X)$.

解: 由于 X 表示 n 次独立重复试验的成功次数, 而每次成功的概率为 p, X 可表示成: $X = X_1 + \cdots + X_n$, 其中 X_i 为独立同分布的伯努利随机变量

$$X_i = \begin{cases} 1 & \text{第 } i \text{ 次试验成功} \\ 0 & \text{其他} \end{cases}$$

利用 (4.1) 可得 $\text{Var}(X) = \text{Var}(X_1) + \cdots + \text{Var}(X_n)$. 又有,

$$\begin{aligned}\text{Var}(X_i) &= E[X_i^2] - (E[X_i])^2 = E[X_i] - (E[X_i])^2 \qquad \text{由于} X_i^2 = X_i \\ &= p - p^2\end{aligned}$$

因此, $\text{Var}(X) = np(1-p)$ ■

例 4c (有限总体中的抽样) 设一共有 N 个人, 每一个人对某件事情有一个态

度, 用一个实数 v 表示一个人对该事件的"支持态度". 用 v_i 表示第 i 个人的支持态度, $i=1,2,\cdots,N$.

现假定一个统计工作者为了获得关于 v_i 的信息, 他"随机地选定" n 个人, 了解他们的 v 值. 所谓随机地选定 n 个人, 是指所有 n 个人的组都是等可能地被选定, 而这种组一共有 $\binom{N}{n}$ 个. 当这 n 个人被选定之后, 就征询他们对某事件的态度并确定他们的 v 值. 用 S 表示这些被选定的 n 个人的 v 值的总和, 求它的期望和方差.

上面提到的问题的一个重要的特例是, 在未来的选举中, 某社区中的每个人对于某项建议或者候选人有一态度, 或者支持或者反对, 若取 v_i 为 1 表示第 i 个人支持, 取 0 为反对, 则 $\bar{v}=\sum_{i=1}^{N} v_i/N$ 表示社区中表示支持的人数比例. 为了估计 $\bar{v}$, 取一个 n 个人的随机样本, 让他们进行投票表态, 样本中表示支持的人数所占的比例 S/n, 通常用作 $\bar{v}$ 的估计.

解: 对每一个人 $i, i=1,\cdots,N$, 确定一个示性函数 I_i, 表示这个人是否在样本中,

$$I_i=\begin{cases}1 & i\ \text{在随机样本中}\\ 0 & \text{其他}\end{cases}$$

S 具有下列表达式

$$S=\sum_{i=1}^{N} v_i I_i$$

故

$$\begin{aligned}E[S]&=\sum_{i=1}^{N} v_i E[I_i]\\ \operatorname{Var}(S)&=\sum_{i=1}^{N}\operatorname{Var}(v_i I_i)+2\sum_{i<j}\sum\operatorname{Cov}(v_i I_i, v_j I_j)\\ &=\sum_{i=1}^{N} v_i^2\operatorname{Var}(I_i)+2\sum_{i<j}\sum v_i v_j\operatorname{Cov}(I_i, I_j)\end{aligned}$$

由于

$$E[I_i]=\frac{n}{N},\quad E[I_i I_j]=\frac{n(n-1)}{N(N-1)}$$

可得

$$\begin{aligned}\operatorname{Var}(I_i)&=\frac{n}{N}\left(1-\frac{n}{N}\right)\\ \operatorname{Cov}(I_i, I_j)&=\frac{n(n-1)}{N(N-1)}-\left(\frac{n}{N}\right)^2=-\frac{n(N-n)}{N^2(N-1)}\end{aligned}$$

因此,

$$E[S] = n\sum_{i=1}^{N}\frac{v_i}{N} = n\bar{v}$$

$$\mathrm{Var}(S) = \frac{n}{N}\left(\frac{N-n}{N}\right)\sum_{i=1}^{N}v_i^2 - \frac{2n(N-n)}{N^2(N-1)}\sum\sum_{i<j}v_iv_j$$

利用恒等式 $(v_1+\cdots+v_N)^2 = \sum_{i=1}^{N}v_i^2 + 2\sum\sum_{i<j}v_iv_j$, $\mathrm{Var}(S)$ 可简化为

$$\mathrm{Var}(S) = \frac{n(N-n)}{N-1}\left(\frac{\sum_{i=1}^{N}v_i^2}{N} - \bar{v}^2\right)$$

现在考虑 Np 个 v_i 为 1, 其余的 v_i 值为 0 的特殊情况. 此时, S 是一个超几何随机变量, 其期望和方差为

$$E[S] = n\bar{v} = np \qquad 由于\bar{v} = \frac{Np}{N} = p$$

$$\mathrm{Var}(S) = \frac{n(N-n)}{N-1}\left(\frac{Np}{N} - p^2\right) = \frac{n(N-n)}{N-1}p(1-p)$$

S/n 表示样本中 v_i 取值为 1 的那一部分的比例. S/n 的方差和期望如下:

$$E\left[\frac{S}{n}\right] = p, \quad \mathrm{Var}\left(\frac{S}{n}\right) = \frac{N-n}{n(N-1)}p(1-p)$$ ■

设 X, Y 为两个随机变量, 假定 $\mathrm{Var}(X)$ 和 $\mathrm{Var}(Y)$ 均大于 0, 则

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

称为 X 和 Y 的**相关系数**, 可以证明

$$-1 \leqslant \rho(X,Y) \leqslant 1 \tag{4.2}$$

为证明 (4.2) 式, 令 $\sigma_x^2 = \mathrm{Var}(X), \sigma_y^2 = \mathrm{Var}(Y)$. 利用不等式

$$0 \leqslant \mathrm{Var}\left(\frac{X}{\sigma_x} + \frac{Y}{\sigma_y}\right) = \frac{\mathrm{Var}(X)}{\sigma_x^2} + \frac{\mathrm{Var}(Y)}{\sigma_y^2} + \frac{2\mathrm{Cov}(X,Y)}{\sigma_x\sigma_y} = 2[1+\rho(X,Y)]$$

可知 $-1 \leqslant \rho(X,Y)$. 另一方面, 由不等式

$$0 \leqslant \mathrm{Var}\left(\frac{X}{\sigma_x} - \frac{Y}{\sigma_y}\right) = \frac{\mathrm{Var}(X)}{\sigma_x^2} + \frac{\mathrm{Var}(Y)}{\sigma_y^2} - \frac{2\mathrm{Cov}(X,Y)}{\sigma_x\sigma_y} = 2[1-\rho(X,Y)]$$

可知 $\rho(X,Y) \leqslant 1$. 因此, 不等式 (4.2) 成立.

事实上, 由 $\mathrm{Var}(Z) = 0$ 可推知随机变量 Z 以概率为 1 地等于一个常数 (第 8 章将严格地证明这一事实). 由 (4.2) 式的证明可以看出, 若 $\rho(X,Y) = 1$, 可推导得

$Y=a+bX$, 其中 $b=\sigma_y/\sigma_x>0$, 同理, 若 $\rho(X,Y)=-1$, 可推导得 $Y=a+bX$, 其中 $b=-\sigma_y/\sigma_x<0$. 现在我们将下列逆命题作为练习留给读者: 若 $Y=a+bX$, 则 $\rho(X,Y)=+1$或-1, 其正负号由 b 的正负所决定.

相关系数是 X,Y 之间线性依赖程度的一种度量. 当 $\rho(X,Y)$ 接近 $+1$ 或 -1 时, 表明 X 与 Y 之间具有很高的线性依赖性, 而当 $\rho(X,Y)$ 接近 0 时, 表示两者之间缺乏这种线性依赖性. 当 $\rho(X,Y)$ 取正值时, 说明当 X 增加时, Y 趋于增加, 而负值说明当 X 增加时, Y 趋于下降. 若 $\rho(X,Y)=0$, 则称 X,Y 为**不相关**(uncorrelated).

例 4d 记 I_A,I_B 为事件 A,B 的示性函数, 即

$$I_A=\begin{cases}1 & A\text{ 发生}\\0 & \text{其他}\end{cases}\qquad I_B=\begin{cases}1 & B\text{ 发生}\\0 & \text{其他}\end{cases}$$

则

$$E[I_A]=P(A),\quad E[I_B]=P(B),\quad E[I_AI_B]=P(AB)$$

故

$$\text{Cov}(I_A,I_B)=P(AB)-P(A)P(B)=P(B)[P(A|B)-P(A)]$$

由上式可得到一个非常直观的结论: I_A 和 I_B 为正相关, 不相关或负相关, 只需看 $P(A|B)$ 是大于、等于或小于 $P(A)$ 即可. ■

下面的例子指出样本均值与离差是不相关的.

例 4e 设 $X_1,\cdots,X_n$ 为独立同分布序列, 其公共方差为 σ^2, 指出

$$\text{Cov}(X_i-\bar{X},\bar{X})=0$$

解:

$$\text{Cov}(X_i-\bar{X},\bar{X})=\text{Cov}(X_i,\bar{X})-\text{Cov}(\bar{X},\bar{X})=\text{Cov}\left(X_i,\frac{1}{n}\sum_{j=1}^{n}X_j\right)-\text{Var}(\bar{X})$$

$$=\frac{1}{n}\sum_{j=1}^{n}\text{Cov}(X_i,X_j)-\frac{\sigma^2}{n}=\frac{\sigma^2}{n}-\frac{\sigma^2}{n}=0$$

其中第三个等式是利用了例 4a 的结果, 最后的等式是由于下列等式的结果,

$$\text{Cov}(X_i,X_j)=\begin{cases}0 & i\neq j & \text{由独立性}\\ \sigma^2 & i=j & \text{由}\text{Var}(X_i)=\sigma^2\end{cases}$$

尽管 $\bar{X}$ 与 $X_i-\bar{X}$ 不相关, 但通常它们并不相互独立. 然而当 X_i 为正态分布时, $\bar{X}$ 不仅与 $X_i-\bar{X}$ 独立, 而且与整个序列 $X_j-\bar{X},j=1,2,\cdots,n$ 相互独立, 这些结果将在 7.8 节给出. 在 7.8 节中还将指出在正态的假定之下, $\bar{X}$ 与样本方差 S^2 也相互独立, 并且 $(n-1)S^2/\sigma^2$ 具有自由度为 $(n-1)$ 的 χ^2 分布 (关于 S^2 之定义

见例 4a). ■

例 4f　考虑 m 个独立实验, 每个实验具有 r 个可能的实验结果, 相应出现的概率分别为 $P_1, \cdots, P_r, \sum_{i=1}^{r} P_i = 1$. 令 N_i 表示 m 次实验中结果 i 出现的次数, 则 $N_1, \cdots, N_r$ 具有多项分布

$$P\{N_1=n_1, N_2=n_2, \cdots, N_r=n_r\} = \frac{m!}{n_1!n_2!\cdots n_r!} P_1^{n_1} P_2^{n_2} \cdots P_r^{n_r} \qquad \sum_{i=1}^{r} n_i = m$$

对于 $i \neq j$, 当 N_i 增大时, N_j 应趋向于变小, 因此, 直观地看来, N_i 与 N_j 应为负相关. 现在计算它们的协方差. N_i 和 N_j 具有下列表达式

$$N_i = \sum_{k=1}^{m} I_i(k) \quad \text{和} \quad N_j = \sum_{k=1}^{m} I_j(k)$$

其中

$$I_i(k) = \begin{cases} 1 & \text{第 } k \text{ 次试验的结果为} i \\ 0 & \text{其他} \end{cases} \qquad I_j(k) = \begin{cases} 1 & \text{第 } k \text{ 次试验的结果为} j \\ 0 & \text{其他} \end{cases}$$

利用命题 4.2(iv), 得到 $\text{Cov}(N_i, N_j) = \sum_{\ell=1}^{m} \sum_{k=1}^{m} \text{Cov}(I_i(k), I_j(\ell))$. 由于第 k 次试验结果是与第 ℓ 次试验结果是相互独立的, 可知

$$\text{Cov}(I_i(k), I_j(\ell)) = 0 \qquad k \neq \ell$$

另一方面,

$$\text{Cov}(I_i(\ell), I_j(\ell)) = E[I_i(\ell)I_j(\ell)] - E[I_i(\ell)]E[I_j(\ell)] = 0 - P_iP_j = -P_iP_j$$

这样, 我们得到

$$\text{Cov}(N_i, N_j) = -mP_iP_j \qquad i \neq j$$

由上式可知 N_i 和 N_j 是负相关的, 符合我们的直观看法. ■

7.5　条件期望

7.5.1　定义

当 X 和 Y 的联合分布为离散分布时, 对于 $P\{Y=y\}>0$ 的 y 值, 给定 $Y=y$ 之下, X 的条件分布列由下式定义:

$$p_{X|Y}(x|y) = P\{X=x|Y=y\} = \frac{p(x,y)}{p_Y(y)}$$

很自然地, 对于 $p_Y(y)>0$, X 在给定 $Y=y$ 之下的条件期望由下式给出

$$E[X|Y=y] = \sum_x xP\{X=x|Y=y\} = \sum_x xp_{X|Y}(x|y)$$

例 5a 设 X 和 Y 独立同分布, 其公共分布为二项分布, 其参数为 (n,p). 计算在 $X+Y=m$ 的条件下 X 的条件期望.

解: 首先计算 $X+Y=m$ 之下, X 的条件分布列, 对于 $k\leqslant \min(n,m)$,

$$P\{X=k|X+Y=m\}=\frac{P\{X=k,X+Y=m\}}{P\{X+Y=m\}}$$
$$=\frac{P\{X=k,Y=m-k\}}{P\{X+Y=m\}}=\frac{P\{X=k\}P\{Y=m-k\}}{P\{X+Y=m\}}$$
$$=\frac{\binom{n}{k}p^k(1-p)^{n-k}\binom{n}{m-k}p^{m-k}(1-p)^{n-m+k}}{\binom{2n}{m}p^m(1-p)^{2n-m}}=\frac{\binom{n}{k}\binom{n}{m-k}}{\binom{2n}{m}}$$

此处, 我们利用了 $X+Y$ 的分布列为二项分布 [参数为 $(2n,p)$]. 因此, 在 $X+Y=m$ 的条件下, X 的分布为超几何分布. 由例 2g, 我们得到

$$E[X|X+Y=m]=m/2$$ ■

类似地, 设 X 和 Y 的联合分布连续, 其联合密度函数为 $f(x,y)$, 对于给定的 $Y=y$, 只要 $f_Y(y)>0$, X 的条件密度函数由下式给出:

$$f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}$$

很自然地, 给定 $Y=y$ 的条件下, X 的条件期望由下式给出

$$E[X|Y=y]=\int_{-\infty}^{\infty}xf_{X|Y}(x|y)\mathrm{d}x$$

此处假定 $f_Y(y)>0$.

例 5b 设 X 和 Y 的联合密度函数为

$$f(x,y)=\frac{\mathrm{e}^{-x/y}\mathrm{e}^{-y}}{y}\qquad 0<x<\infty,0<y<\infty$$

计算 $E[X|Y=y]$.

解: 先计算条件密度

$$f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}=\frac{f(x,y)}{\int_{-\infty}^{\infty}f(x,y)\mathrm{d}x}=\frac{(1/y)\mathrm{e}^{-x/y}\mathrm{e}^{-y}}{\int_0^{\infty}(1/y)\mathrm{e}^{-x/y}\mathrm{e}^{-y}\mathrm{d}x}$$
$$=\frac{(1/y)\mathrm{e}^{-x/y}}{\int_0^{\infty}(1/y)\mathrm{e}^{-x/y}\mathrm{d}x}=\frac{1}{y}\mathrm{e}^{-x/y}$$

因此, X 在给定 $Y=y$ 之下的条件分布刚好是指数分布, 其期望为 y, 因此

$$E[X|Y=y]=\int_0^{\infty}\frac{x}{y}\mathrm{e}^{-x/y}\mathrm{d}x=y$$ ■

注释 条件概率满足概率的所有性质, 条件期望也满足通常期望的性质, 例如

公式

$$E[g(X)|Y=y]=\begin{cases}\sum_x g(x)p_{X|Y}(x|y) & \text{离散情形}\\ \int_{-\infty}^{\infty} g(x)f_{X|Y}(x|y)\,\mathrm{d}x & \text{连续情形}\end{cases}$$

$$E\Big[\sum_{i=1}^{n} X_i|Y=y\Big]=\sum_{i=1}^{n} E[X_i|Y=y]$$

仍然有效. 事实上, 给定 $Y=y$ 之下的条件期望可以看成是较小的样本空间中的普通的期望, 这个小的样本空间由满足条件 $\{Y=y\}$ 的那些样本点组成. ■

7.5.2 利用条件计算期望

用 $E[X|Y]$ 表示随机变量 Y 的函数, 它在 $Y=y$ 处的值为 $E[X|Y=y]$, 注意 $E[X|Y]$ 本身是一个随机变量. 下面给出的命题是条件期望一个极其重要的性质.

命题 5.1

$$E[X]=E[E[X|Y]] \tag{5.1}$$

若 Y 是离散型随机变量, (5.1) 变成

$$E[X]=\sum_y E[X|Y=y]P\{Y=y\} \tag{5.1a}$$

若 Y 的分布连续, 具有密度函数 $f_Y(y)$, 则 (5.1) 式变成

$$E[X]=\int_{-\infty}^{\infty} E[X|Y=y]f_Y(y)\mathrm{d}y \tag{5.1b}$$

X 和 Y 为离散情形下 (5.1) 式的证明: 我们必须指出

$$E[X]=\sum_y E[X|Y=y]P\{Y=y\} \tag{5.2}$$

从等式右边开始,

$$\begin{aligned}\sum_y E[X|Y=y]P\{Y=y\} &= \sum_y\sum_x xP\{X=x|Y=y\}P\{Y=y\}\\ &=\sum_y\sum_x x\frac{P\{X=x,Y=y\}}{P\{Y=y\}}P\{Y=y\}=\sum_y\sum_x xP\{X=x,Y=y\}\\ &=\sum_x x\sum_y P\{X=x,Y=y\}=\sum_x xP\{X=x\}=E[X]\end{aligned}$$

□

可以这样来解释 (5.2) 式, 期望值 $E[X]$ 可以看成条件期望 $E[X|Y=y]$ 的加权平均, 而权值刚好是条件 $\{Y=y\}$ 的概率. 这个结果对计算随机变量的期望是极其重要的, 它可以让我们首先很容易地计算某随机变量在给定条件之下的条件期望, 然后再对条件期望求平均. 下面的例子说明了这个公式的用处.

例 5c 一个矿工在井下迷了路, 迷路的地方有三个门, 从第一个门出来, 经过 3 个小时后, 可到达安全之处. 若从第二个门出去, 经过 5 个小时后, 他会回到原地. 若从第三个门出来, 经过 7 个小时后才会到原地. 假定工人在任何时候都是随机地选择一个门. 问这个工人为了走到安全之处, 平均需要多少时间.

解: 设 X 表示该矿工为到达安全之处所需的时间, 又设 Y 为他选择的门的号码, 则

$$\begin{aligned}E[X] &= E[X|Y=1]P\{Y=1\}+E[X|Y=2]P\{Y=2\}+E[X|Y=3]P\{Y=3\}\\&=\frac{1}{3}(E[X|Y=1]+E[X|Y=2]+E[X|Y=3]),\end{aligned}$$

然而

$$E[X|Y=1]=3 \quad E[X|Y=2]=5+E[X] \quad E[X|Y=3]=7+E[X] \tag{5.3}$$

我们在此对 (5.3) 式作一解释, 例如 $E[X|Y=2]$ 的公式, 其理由如下: 如选择第二个门, 他花 5 个小时后又回到了原地, 但回到原地, 问题与刚开始时一样, 他到达安全地点所需时间为 $E[X]$. 因此 $E[X|Y=2]=5+E[X]$. (5.3) 的其余公式是完全类似的. 因此

$$E[X]=\frac{1}{3}(3+5+E[X]+7+E[X])$$

从而 $E[X]=15$. ■

例 5d (随机个数随机变量之和的期望) 假设在某一天进入百货商店的人数是一个随机变量, 其平均数为 50. 进一步假定这些顾客在店里花费的钱数是相互独立且同分布的随机变量, 每位顾客的花钱数的期望为 8 元, 并且顾客的花钱数与进入百货店的人数也是相互独立的. 求那一天百货店营业额的期望值.

解: 记 N 为进入百货店的顾客人数, X_i 是顾客 i 在店内的消费额, 则百货店内消费总量为 $\sum_{i=1}^{N} X_i$

$$E\Big[\sum_{i=1}^{N} X_i\Big]=E\Big[E\Big[\sum_{i=1}^{N} X_i|N\Big]\Big]$$

但

$$\begin{aligned}E\Big[\sum_{i=1}^{N} X_i|N=n\Big]&=E\Big[\sum_{i=1}^{n} X_i|N=n\Big]\\&=E\Big[\sum_{i=1}^{n} X_i\Big] \qquad \text{由 } X_i \text{ 与 } N \text{ 的独立性}\\&=nE[X] \qquad \text{其中 } E[X]=E[X_i]\end{aligned}$$

由此可得,

$$E\left[\sum_{i=1}^{N} X_i|N\right]=NE[X]$$

从而

$$E\left[\sum_{i=1}^{N} X_i\right] = E[NE[X]] = E[N]E[X]$$

因此, 当天百货店营业额的期望值为 $50 \times 8 = 400$ 元. ■

例 5e　掷骰子的 "Crap" 游戏是这样的, 每次掷两枚骰子, 开始时, 如果得到的点数之和是 2, 3 或 12 则玩者输, 若得到 7 或 11, 则玩者赢, 若得到的是其他点数 i, 则需继续玩下去, 一直到掷出 7 或 i 为止. 若玩者最后得到的点数为 7, 则玩者输, 若最后得到的为 i 则玩者赢. 记 R 为掷骰子的次数, 求:

(a) $E[R]$;　(b) $E[R|\text{玩者赢}]$;　(c) $E[R|\text{玩者输}]$.

解: 记 P_i 为每次掷骰子得到两个骰子之和为 i 之概率, 易知

$$P_i = P_{14-i} = \frac{i-1}{36} \qquad i = 2, 3, \cdots, 7$$

记 S 为第一次掷出点数, 则

$$E[R] = \sum_{i=2}^{12} E[R|S=i]P_i$$

其中,

$$E[R|S=i] = \begin{cases} 1 & i = 2, 3, 7, 11, 12 \\ 1 + \dfrac{1}{P_i + P_7} & \text{其他} \end{cases}$$

在上式中, 若第一次得到 i, $i \neq 2, 3, 7, 11$或12, 则玩游戏者必须继续进行直到出现 i 或 7 为止, 而掷骰子的次数服从几何分布, 参数为 $P_i + P_7$. 因此, 掷骰子的期望次数为 $\dfrac{1}{P_i + P_7} + 1$, 其中 $+1$ 表示加上第一次掷骰子, 这样,

$$E[R] = 1 + \sum_{i=4}^{6} \frac{P_i}{P_i + P_7} + \sum_{i=8}^{10} \frac{P_i}{P_i + P_7} = 1 + 2(3/9 + 4/10 + 5/11) = 3.376$$

现在计算 $E[R|\text{赢}]$, 为此, 先计算玩游戏者赢的概率 p. 将第一次掷骰子的结果 S 作为条件,

$$p = \sum_{i=2}^{12} P\{\text{赢}|S=i\}P_i = P_7 + P_{11} + \sum_{i=4}^{6} \frac{P_i}{P_i + P_7}P_i + \sum_{i=8}^{10} \frac{P_i}{P_i + P_7}P_i = 0.493$$

上面等式中当 $S = i, i = 4, 5, 6$或$8, 9, 10$ 时, $P\{\text{赢}|S=i\}$ 就是在后续掷骰子过程中, i 在 7 之前出现的概率, 值为 $P_i/(P_i + P_7)$. 现在需要确定玩游戏者赢的条件下 S 的条件概率, 记 $Q_i = P\{S = i|\text{赢}\}$, 我们有

$$Q_2 = Q_3 = Q_{12} = 0 \qquad Q_7 = P_7/p \qquad Q_{11} = P_{11}/p$$

对于 $i = 4, 5, 6, 8, 9, 10$

$$Q_i = \frac{P\{S=i, \text{赢}\}}{P\{\text{赢}\}} = \frac{P_i P\{\text{赢}|S=i\}}{p} = \frac{P_i^2}{p(P_i + P_7)}$$

这样,

$$E[R|\text{赢}] = \sum_i E[R|\text{赢}, S=i]Q_i$$

在第 6 章例 2j 已经指出, 已知 $S=i$ 的条件之下, 需要掷多少次骰子与最后的结果是赢或输是相互独立的.(可以这样来看这个事实, 在需要掷的次数为 R 的条件下, 是赢是输的概率与已经掷了几次是无关的, 再利用事件独立性的对称特性, 即 A 独立于事件 B, 则 B 也独立于事件 A, 可以推出在输赢已知的条件下, R 的分布与输赢也是无关的.) 这样我们得到

$$E[R|\text{赢}] = \sum_i E[R|S=i]Q_i = 1 + \sum_{i=4}^{6} \frac{Q_i}{P_i+P_7} + \sum_{i=8}^{10} \frac{Q_i}{P_i+P_7} = 2.938$$

虽然我们可以仿照计算 $E[R|\text{赢}]$ 一样, 计算 $E[R|\text{输}]$. 但是由下面的公式

$$E[R] = E[R|\text{赢}]p + E[R|\text{输}](1-p)$$

可立即求得

$$E[R|\text{输}] = \frac{E[R]-E[R|\text{赢}]p}{1-p} = 3.801$$ ■

例 5f 在第 6 章例 5c 已经指出, 二维正态随机向量 (X,Y) 的联合密度为

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]\right\}$$

现在需要指出 ρ 是 X 与 Y 的相关系数. 第 6 章例 5c 已经指出, 在 $Y=y$ 之下, X 具有正态分布, 其条件期望为 $\mu_x + \rho\frac{\sigma_x}{\sigma_y}(y-\mu_y)$, 由此得到

$$\begin{aligned} E[XY|Y=y] &= E[Xy|Y=y] = yE[X|Y=y] \\ &= y\left[\mu_x + \rho\frac{\sigma_x}{\sigma_y}(y-\mu_y)\right] = y\mu_x + \rho\frac{\sigma_x}{\sigma_y}(y^2-\mu_y y) \end{aligned}$$

从而

$$E[XY|Y] = Y\mu_x + \rho\frac{\sigma_x}{\sigma_y}(Y^2-\mu_y Y)$$

由上式可得

$$\begin{aligned} E[XY] &= E\left[Y\mu_x + \rho\frac{\sigma_x}{\sigma_y}(Y^2-\mu_y Y)\right] = \mu_x E[Y] + \rho\frac{\sigma_x}{\sigma_y}E[Y^2-\mu_y Y] \\ &= \mu_x\mu_y + \rho\frac{\sigma_x}{\sigma_y}(E[Y^2]-\mu_y^2) = \mu_x\mu_y + \rho\frac{\sigma_x}{\sigma_y}\text{Var}(Y) = \mu_x\mu_y + \rho\sigma_x\sigma_y \end{aligned}$$

由此可得

$$\mathrm{Corr}(X,Y)=\frac{\mathrm{Cov}(X,Y)}{\sigma_x\sigma_y}=\frac{E[XY]-\mu_x\mu_y}{\sigma_x\sigma_y}=\frac{\rho\sigma_x\sigma_y}{\sigma_x\sigma_y}=\rho$$ ■

有时候, 随机变量 X 的期望 $E[X]$ 反而容易计算, 此时我们可利用关于条件期望的恒等式 (即 (5.1a) 或 (5.1b)) 来计算条件期望, 请看下面的例子.

例 5g 考虑 n 次独立重复试验, 每次试验的结果为 $1,2,\cdots,k$, 相应的概率为 $p_1,\cdots,p_k,\sum_{i=1}^k p_i=1$. 令 N_i 表示试验中出现结果 i 的次数. 对于 $i\neq j$, 计算

(a) $E[N_j|N_i>0]$. (b) $E[N_j|N_i>1]$.

解: 对于 (a), 令

$$I=\begin{cases}0, & N_i=0\\ 1, & N_i>0\end{cases}$$

$E[N_j]$ 可以写成

$$E[N_j]=E[N_j|I=0]P\{I=0\}+E[N_j|I=1]P\{I=1\}$$

或等价地

$$E[N_j]=E[N_j|N_i=0]P\{N_i=0\}+E[N_j|N_i>0]P\{N_i>0\}$$

对于 N_j, 其无条件分布是二项分布, 相应的参数为 (n,p_j). 现在设 $N_i=r$ 给定, 则其余的 $N-r$ 次试验不会取到 i, 并且相互独立地取 j 的概率为 $P(j|不是i)=p_j/(1-p_i)$. 因此 N_j 在 $N_i=r$ 之下的条件分布为二项分布, 其参数为 $(n-r,p_j/(1-p_i))$(关于推导的细节可参见第 6 章例 4c). 又由于 $P\{N_i=0\}=(1-p_i)^n$. 因此, 上面关于 $E[N_j]$ 的等式变成

$$np_j=n\frac{p_j}{1-p_i}(1-p_i)^n+E[N_j|N_i>0](1-(1-p_i)^n)$$

上式给出

$$E[N_j|N_i>0]=np_j\frac{1-(1-p_i)^{n-1}}{1-(1-p_i)^n}$$

对于 (b), 其讨论是完全类似的, 令

$$J=\begin{cases}0, & N_i=0\\ 1, & N_i=1\\ 2, & N_i>1\end{cases}$$

得到

$$\begin{aligned}E[N_j]=&E[N_j|J=0]P\{J=0\}+E[N_j|J=1]P\{J=1\}\\&+E[N_j|J=2]P\{J=2\}\end{aligned}$$

或等价地

$$E[N_j] = E[N_j|N_i=0]P\{N_i=0\} + E[N_j|N_i=1]P\{N_i=1\} + E[N_j|N_i>1]P\{N_i>1\}$$

由这个公式导出

$$np_j = n\frac{p_j}{1-p_i}(1-p_i)^n + (n-1)\frac{p_j}{1-p_i}np_i(1-p_i)^{n-1} + E[N_j|N_i>1](1-(1-p_i)^n - np_i(1-p_i)^{n-1})$$

最后得到

$$E[N_j|N_i>1] = \frac{np_j[1-(1-p_i)^{n-1}-(n-1)p_i(1-p_i)^{n-2}]}{1-(1-p_i)^n - np_i(1-p_i)^{n-1}}$$

■

类似地，也可以利用条件的方法计算随机变量的方差. 请看下面的例子:

例 5h (几何分布的方差) 设有一独立重复试验序列，每次试验成功的概率为 p，记 N 为取得第一次成功所需的试验次数. 求 $\mathrm{Var}(N)$.

解: 令 $Y=1$，若第一次试验成功; $Y=0$，其他情况. 利用公式

$$\mathrm{Var}(N) = E[N^2] - (E[N])^2$$

为计算 $E[N^2]$，先将它写成如下形式

$$E[N^2] = E[E[N^2|Y]]$$

由于

$$E[N^2|Y=1] = 1$$

$$E[N^2|Y=0] = E[(1+N)^2]$$

上述两式成立的理由如下，若第一次实验成功，显然 $N=1$，从而 $N^2=1$，故第一式成立. 若 $Y=0$，即第一次试验失败，则在此种情况下试验相当于重新开始，因此为了达到第一次成功所需实验次数变成 $N+1$，我们得到第二式. 这样

$$E[N^2] = E[N^2|Y=1]P\{Y=1\} + E[N^2|Y=0]P\{Y=0\}$$
$$= p + (1-p)E[(1+N)^2] = 1 + (1-p)E[2N+N^2]$$

在第 4 章例 8b 中，已经指出 $E[N]=1/p$，因此我们得到

$$E[N^2] = 1 + \frac{2(1-p)}{p} + (1-p)E[N^2]$$

由上式解得

$$E[N^2] = \frac{2-p}{p^2}$$

从而

$$\mathrm{Var}(N) = E[N^2] - (E[N])^2 = \frac{2-p}{p^2} - \left(\frac{1}{p}\right)^2 = \frac{1-p}{p^2}$$

■

例 5i 考虑有 r 个玩家的赌博问题, 开始时玩家 i 具有 n_i 个单位的赌资, $n_i > 0, i = 1, \cdots, r$. 每一盘, 选出两人进行赌博, 胜者从败者拿走一个单位赌资, 任何人, 只要赌资变成 0, 就退出比赛. 一直到某个人赢得所有赌资 $n \equiv \sum_{i=1}^{r} n_i$ 个单位. 这个人就是最后的赢家. 假定各盘的输赢都是相互独立的. 并且参加赌博的任何一对玩家, 两人都有相同的赢的概率. 求结束赌博所需盘数的期望.

解: 首先, 我们假定只有两个玩家. 玩家 1 和玩家 2 的最初的赌资分别为 j 和 $n-j$ 个单位. 记 X_j 为结束赌博所需的盘数. 记 $m_j = E[X_j]$, 对于 $j = 1, \cdots, n-1$,

$$X_j = 1 + A_j$$

其中 A_j 表示第一盘以后所需的附加盘数. 由此得到

$$m_j = 1 + E[A_j]$$

以第一盘的结果为条件, 可得

$$m_j = 1 + E[A_j|1\text{ 赢第一盘}] \cdot \frac{1}{2} + E[A_j|2\text{ 赢第一盘}] \cdot \frac{1}{2}$$

如果玩家 1 赢, 此时玩家 1 的赌资变成 $j+1$, 而玩家 2 的赌资变成 $n-j-1$, 显然

$$E[A_j|1\text{赢第一盘}] = m_{j+1}, \quad E[A_j|2\text{ 赢第一盘}] = m_{j-1}$$

这样, 我们得到

$$m_j = 1 + \frac{1}{2}m_{j+1} + \frac{1}{2}m_{j-1}$$

或

$$m_{j+1} = 2m_j - m_{j-1} - 2, \quad j = 1, \cdots, n-1 \tag{5.4}$$

利用 $m_0 = 0$, 得

$$\begin{aligned}
m_2 &= 2m_1 - 2 \\
m_3 &= 2m_2 - m_1 - 2 = 3m_1 - 6 = 3(m_1 - 2) \\
m_4 &= 2m_3 - m_2 - 2 = 4m_1 - 12 = 4(m_1 - 3)
\end{aligned}$$

由此, 看出有如下规律,

$$m_i = i(m_1 - i + 1), \quad i = 1, 2, \cdots, n \tag{5.5}$$

要证明上式, 只需利用数学归纳法. 我们已经证明对于 $i = 1, 2$, (5.5) 式成立, 假定 (5.5) 式对一切 $i \leqslant j < n$ 成立, 利用 (5.4),

$$\begin{aligned}
m_{j+1} &= 2m_j - m_{j-1} - 2 \\
&= 2j(m_1 - j + 1) - (j-1)(m_1 - j + 2) - 2 \qquad (\text{由归纳法假设}) \\
&= (j+1)m_1 - 2j^2 + 2j + j^2 - 3j + 2 - 2 \\
&= (j+1)m_1 - j^2 - j = (j+1)(m_1 - j)
\end{aligned}$$

由归纳法, 我们证明了 (5.5) 式, 利用 $m_n = 0$, 可得

$$m_1 = n - 1$$

再利用 (5.5) 式, 可得

$$m_i = i(n-i)$$

这样, 当只有两个玩家且初始赌资分别为 i 与 $n-i$ 的时候, 所玩的盘数的均值为 $i(n-i)$, 这个数也是玩家 1 玩的平均盘数.

现在回到 r 个玩家参加赌博的游戏, r 个玩家每人开始时, 手中的赌资为 n_i, $i = 1, \cdots, r, \sum_{i=1}^r n_i = n$. 令 X 表示从赌博一开始到结束所赌的盘数. 令 X_i 为玩家 i 自始至终所赌的盘数. 虽然, 在 i 参加的各盘赌博中, 其对手不一. 由于各对手都是势均力敌. 对于 i 来说, 他的各盘赌博就像是与一个"人"赌似的, 但这个人的赌资是 $n - n_i$. 这样, 第 i 人参赌的盘数的平均值为 $E[X_i] = n_i(n - n_i)$. 这 r 个人参赌的总盘数的平均值为

$$E\left[\sum_{i=1}^r X_i\right] = \sum_{i=1}^r n_i(n - n_i) = n^2 - \sum_{i=1}^r n_i^2$$

但实际赌博时, 都是两人参加一盘赌博. 因此赌博从开始一直到结束的盘数 $X = \frac{1}{2}\sum_{i=1}^r X_i$. 这样,

$$E[X] = \frac{1}{2}\left(n^2 - \sum_{i=1}^r n_i^2\right)$$

十分有意义的是, 我们的论证显示, 从开始到结束, 赌博的平均盘数并不依赖于对手的选择. 但是, X 的分布却是依赖于对手的选择. 例如, $r=3$, $n_1 = n_2 = 1$, $n_3 = 2$. 如果第一盘让赌徒 1 和赌徒 2 先赌. 那么至少要赌 3 盘才能决出最后的胜者, 即 $P\{X \geqslant 3\} = 1$. 但是让赌徒 3 在第一盘中玩, 那么就有可能在 2 盘之内决出最后的胜利者. 即 $P\{X = 2\} > 0$. 这说明 X 的分布是依赖于对手的选择的. ■

下面的例子中, 我们利用条件期望的性质验证 6.3.1 节中的一个结果: (0 ,1) 上独立同分布均匀随机变量序列的部分和最早超过 1 的部分和项数的期望值竟然等于 e.

例 5j 设 $U_1, U_2, \cdots$ 为一相互独立的 $(0,1)$ 上均匀随机变量序列, 求 $E[N]$, 其中

$$N = \min\left\{n : \sum_{i=1}^n U_i > 1\right\}$$

解: 我们将得到一个更一般的结果. 对于 $x \in [0,1]$, 令

$$N(x) = \min\left\{n : \sum_{i=1}^n U_i > x\right\}$$

及

$$m(x) = E[N(x)]$$

$N(x)$ 是部分和 $\sum_{i=1}^{n} U_i$ 超过 x 的最小指标 n, $m(x)$ 是 $N(x)$ 的期望值. 将 U_1 作为条件, 利用公式 (5.1b), 得到

$$m(x) = \int_0^1 E[N(x)|U_1 = y]\mathrm{d}y \tag{5.6}$$

对于条件期望 $E[N(x)|U_1 = y]$, 我们有

$$E[N(x)|U_1 = y] = \begin{cases} 1, & y > x \\ 1 + m(x - y), & y \leqslant x \end{cases} \tag{5.7}$$

上式中, 当 $y > x$ 的情况等式是很显然的, 当 $y \leqslant x$ 时, 此时需要继续取 $U_2, \cdots$, 此时相当于序列从 U_2 开始, 要求出超过 $x - y$ 的最小时刻. 因此, 这个时刻的期望值应为 $m(x-y)$, 加上原有的 U_1 这个随机数, $E[N(x)|U_1 = y]$ 就变成 (5.5) 的第二种情况. 将 (5.7) 式代入 (5.6) 式, 得到

$$\begin{aligned} m(x) &= 1 + \int_0^x m(x-y)\mathrm{d}y \\ &= 1 + \int_0^x m(u)\mathrm{d}u \qquad (\text{作变量替换 } u = x - y) \end{aligned}$$

将上式求微商得到

$$m'(x) = m(x)$$

或

$$\frac{m'(x)}{m(x)} = 1$$

将上式求积分, 得

$$\ln[m(x)] = x + c$$

或

$$m(x) = k\mathrm{e}^x$$

由 $m(0) = 1$, 得 $k = 1$, 这样

$$m(x) = \mathrm{e}^x$$

回到我们最早提出的问题, 均匀随机数部分和超过 1 的最早时刻的期望值 $m(1)$ 等于 e. ■

7.5.3 利用条件计算概率

我们不仅可以利用条件期望的方法计算一个随机变量的期望, 也可利用此方法计算概率. 设 E 为一随机事件, 令 X 为 E 的示性函数, 即

$$X = \begin{cases} 1 & E\text{发生} \\ 0 & E\text{不发生} \end{cases}$$

由此,

$$E[X] = P(E)$$

$$E[X|Y=y] = P(E|Y=y) \qquad \text{对任何随机变量 } Y$$

这样, 利用 (5.1a) 与 (5.1b), 可得到

$$P(E) = \begin{cases} \sum_y P(E|Y=y)P(Y=y) & Y \text{ 为离散型随机变量} \\ \int_{-\infty}^{\infty} P(E|Y=y)f_Y(y)\mathrm{d}y & Y \text{ 为连续型随机变量} \end{cases} \tag{5.8}$$

若 Y 是离散型随机变量, 并且只取值 $y_1, \cdots, y_n$, 利用记号 $F_i = \{Y = y_i\}$, 方程 (5.6) 变成

$$P(E) = \sum_{i=1}^{n} P(E|F_i)P(F_i)$$

其中 $F_1, \cdots, F_n$ 互不相容, 且这些事件的和是一个必然事件, 此时我们称它们形成一个完备事件组. 这个公式就是概率的全概公式.

例 5k (最优奖问题) 设有 n 个不同的奖陆续出台, 当一个奖出来时, 你可以拒绝或接受. 当然, 你接受了这个刚出台的奖, 你不能再领以后出台的奖. 若你拒绝刚出台的奖, 那么你还有机会领以后出台的奖. 当一个奖出台时, 唯一的信息是刚出台的奖与已经出台的奖进行比较. 例如, 当第 5 个奖出台时, 你只能与前 4 个已经公布的奖进行比较. 我们的目标是希望得到最高奖, 或找到一种策略使得得到最高奖的概率尽可能大. 假设出台的奖项的 $n!$ 种次序都是等可能的.

解: 令人惊讶的是, 我们可以得到很好的结果. 对于固定的 $k, 0 \leqslant k < n$, 考虑如下的策略. 首先拒绝前面 k 个奖项, 然后从第 $k+1$ 个奖项出台开始算起, 只要发现新出台的奖项比前面已经发布的好就接受这个奖项, 否则就拒绝这个奖项而观察出台的下一个奖项, 记 P_k(最优) 表示利用这个策略得到最优奖项的概率, 记 X 为最优奖项出台的次序, 例如最优奖是第 5 个出台的奖, 则 $X = 5$. 利用全概公式

$$P_k(\text{最优}) = \sum_{i=1}^{n} P_k(\text{ 最优}|X=i)P(X=i) = \frac{1}{n}\sum_{i=1}^{n} P_k(\text{最优}|X=i)$$

若最优的奖项在前面的 k 次发布, 按这个选奖的策略, 每次都拒绝拿奖, 因此, 不可能拿到最优奖. 这样

$$P_k(\text{最优}|X=i) = 0 \qquad i \leqslant k$$

另一方面, 若最优奖的位置 i 在 k 之后, 即 $i > k$, 那么就有可能拿到最优奖. 如果前面 $i-1$ 个奖项的最大值奖的位置在前面的 k 个奖中, 那么, 随着奖的出台, 一直到 $i-1$ 都是拒绝领奖, 直到最优奖 i 发布时, 按规则接受最优奖. 现在假定最优奖

位置在 i, 在前面 $i-1$ 个奖项中, 最高奖的位置在 $1,\cdots,i-1$ 处是等可能的. 因此

$$\begin{aligned}P_k(\text{最优}|X=i) &= P\{\text{前面}i-1\text{个奖中, 最高的奖项在}\{1,2,\cdots,k\}\text{中}|X=i\}\\ &= \frac{k}{i-1}, \qquad i>k\end{aligned}$$

这样, 我们得到

$$P_k(\text{ 最优}) = \frac{k}{n}\sum_{i=k+1}^{n}\frac{1}{i-1} \approx \frac{k}{n}\int_{k+1}^{n}\frac{1}{x-1}\mathrm{d}x = \frac{k}{n}\ln\left(\frac{n-1}{k}\right) \approx \frac{k}{n}\ln\left(\frac{n}{k}\right)$$

若考虑函数

$$g(x) = \frac{x}{n}\ln\left(\frac{n}{x}\right)$$

则

$$g'(x) = \frac{1}{n}\ln\left(\frac{n}{x}\right) - \frac{1}{n}$$

因此

$$g'(x) = 0 \Rightarrow \ln\left(\frac{n}{x}\right) = 1 \Rightarrow x = \frac{n}{\mathrm{e}}$$

由于 $P_k(\text{ 最优}) \approx g(k)$, 当取 $k=n/\mathrm{e}$ 时, $P_k(\text{ 最优}) \approx g(n/\mathrm{e}) = 1/\mathrm{e}$, 最优策略是首先拒绝前面 $k=n/\mathrm{e}$ 个奖项, 然后等待出现第一个比以前的奖项都大的奖项, 并接受这个奖项. 按这个策略, 拿到最优奖的概率近似地等于 $1/\mathrm{e} \approx 0.367\,88$. ■

注释 大部分人对于以这么大的概率拿到最优奖感到吃惊. 一般认为这个概率当 n 很大时会趋于 0. 然而, 即使不经精确计算, 稍微思考, 就会发现拿到最优奖的概率会相当大. 取 $k=n/2$, 考虑这 n 个奖中的最高奖与第二最高奖. 考虑一个随机事件: 第二最高奖出现在前面一半, 第一最高奖出现在后面一半. 这个事件的概率为 1/4. 当这个事件发生时, 我们一定能选到奖, 并且是最高奖. 因此看出, n 无论怎么大, 总是能找到一种策略, 使得得到最高奖的可能性超过 1/4. ■

例 51 设 U 为 $(0,1)$ 上均匀分布的随机变量, 又设在给定 $U=p$ 的条件下, 随机变量 X 的分布为二项分布, 其参数为 (n,p). 计算 X 的分布列.

解: 利用 X 的条件分布, 可以计算 X 的分布列,

$$\begin{aligned}P\{X=i\} &= \int_0^1 P\{X=i|U=p\}f_U(p)\mathrm{d}p = \int_0^1 P\{X=i|U=p\}\mathrm{d}p\\ &= \frac{n!}{i!(n-i)!}\int_0^1 p^i(1-p)^{n-i}\mathrm{d}p \qquad i=0,1,\cdots,n\end{aligned}$$

又已知

$$\int_0^1 p^i(1-p)^{n-i}\mathrm{d}p = \frac{i!(n-i)!}{(n+1)!}$$

(这个公式的概率证明可参见 6.6 节), 由此可得

$$P\{X=i\} = \frac{1}{n+1} \qquad i=0,\cdots,n$$

由这个公式, 我们可以得到一个令人吃惊的事实. 如果将一个硬币连续掷 n 次, 假定硬币的正面朝上的概率 p 的分布为 $(0,1)$ 上均匀分布, 则正面朝上的次数的分布是 $\{0,1,\cdots,n\}$ 上的均匀分布.

由于条件分布具有很好的形式, 我们可以给出一个较为直观的解释, 设 U, $U_1,\cdots,U_n$ 为 $n+1$ 个独立同分布的随机变量, 每一个变量的分布为 $(0,1)$ 上的均匀分布. 令 X 为 $U_1,\cdots,U_n$ 中小于 U 的变量个数, 由于 $U_1,\cdots,U_n$ 和 U 具有相同分布, 在 $n+1$ 个变量的排序过程中 U 为最小, U 为第 2 小, $\cdots$, U 为最大, 这 $n+1$ 种可能性是相同的, 因此 X 等于 $0,1,2,3,\cdots,n$ 这 $n+1$ 种可能性也是相同的. 又由于当给定 $U=p$ 的条件下, $U_i \leqslant U, i=1,\cdots,n$ 的个数的分布为二项分布, 其参数为 (n,p). 这样, X 的分布具有很直观的解释. ■

例 5m 设 X 和 Y 为两个相互独立的随机变量, 其密度分别为 f_X 和 f_Y. 计算 $P\{X<Y\}$.

解: 先将 Y 的值固定, 利用公式 (5.8), 得到

$$\begin{aligned} P\{X<Y\} &= \int_{-\infty}^{\infty} P\{X<Y|Y=y\} f_Y(y)\mathrm{d}y \\ &= \int_{-\infty}^{\infty} P\{X<y|Y=y\} f_Y(y)\mathrm{d}y \\ &= \int_{-\infty}^{\infty} P\{X<y\} f_Y(y)\mathrm{d}y \qquad \text{独立性} \\ &= \int_{-\infty}^{\infty} F_X(y) f_Y(y)\mathrm{d}y \end{aligned}$$

其中

$$F_X(y) = \int_{-\infty}^{y} f_X(x)\mathrm{d}x$$ ■

例 5n 设 X, Y 为相互独立的连续随机变量, 其密度分别为 $f_X(x), f_Y(y)$. 求 $X+Y$ 的分布.

解: 利用公式 (5.8), 我们得到

$$\begin{aligned} P\{X+Y<a\} &= \int_{-\infty}^{\infty} P\{X+Y<a|Y=y\} f_Y(y)\mathrm{d}y \\ &= \int_{-\infty}^{\infty} P\{X+y<a|Y=y\} f_Y(y)\mathrm{d}y \\ &= \int_{-\infty}^{\infty} P\{X<a-y\} f_Y(y)\mathrm{d}y = \int_{-\infty}^{\infty} F_X(a-y) f_Y(y)\mathrm{d}y \end{aligned}$$ ■

7.5.4 条件方差

我们既然可以定义 $Y=y$ 之下 X 的条件期望, 也可以定义 $Y=y$ 之下 X 的

条件方差, 它由下式定义

$$\mathrm{Var}(X|Y) \equiv E[(X - E[X|Y])^2|Y]$$

$\mathrm{Var}(X|Y)$ 是 X 和它的条件期望之差的平方的 (条件) 期望值. 或者, 换句话说, $\mathrm{Var}(X|Y)$ 在 Y 已知的条件下与通常的方差的定义完全一样, 不过求期望的过程换成了求条件期望.

条件方差和方差之间具有某种联系, 人们通常利用这种关系计算一个随机变量的方差. 首先, 和普通方差的公式 $\mathrm{Var}(X) = E[X^2] - (E[X])^2$ 一样, 条件方差也有

$$\mathrm{Var}(X|Y) = E[X^2|Y] - (E[X|Y])^2$$

由此得到

$$E[\mathrm{Var}(X|Y)] = E[E[X^2|Y]] - E[(E[X|Y])^2] = E[X^2] - E[(E[X|Y])^2] \tag{5.9}$$

同时,

$$\mathrm{Var}(E[X|Y]) = E[(E[X|Y])^2] - (E[E[X|Y]])^2 = E[(E[X|Y])^2] - (E[X])^2 \tag{5.10}$$

将 (5.9) 式与 (5.10) 式相加, 我们得到如下命题:

命题 5.2 (条件方差公式)

$$\mathrm{Var}(X) = E[\mathrm{Var}(X|Y)] + \mathrm{Var}(E[X|Y])$$

例 5o　设在任何长度为 t 一段时间区间 $(0,t)$ 内到达某火车站的人数是一个服从泊松分布的随机变量, 均值为 λt. 现设火车在 $(0,T)$ 这个区间内随机到达, 即到达时间是 $(0,T)$ 上的均匀分布, 并且它与到达火车站的人数独立. 求火车到达时上火车的旅客人数的期望和方差.

解: 记 $N(t)$ 表示 t 以前到达车站的人数, Y 表示火车到达时间, 能够上火车的人数为 $N(Y)$ (假定火车到达后立即开走, 而 Y 以后到达火车站的人只好等下一班火车). 把 Y 的值固定, 设为条件, 得

$$\begin{aligned} E[N(Y)|Y=t] &= E[N(t)|Y=t] = E[N(t)] && \text{由 } Y \text{ 与 } N(t) \text{ 的独立性} \\ &= \lambda t && N(t) \text{ 是泊松随机变量, 平均值为 } \lambda t \end{aligned}$$

因此,

$$E[N(Y)|Y] = \lambda Y$$

两边求期望, 得到

$$E[N(Y)] = \lambda E[Y] = \frac{\lambda T}{2}$$

为求 $\mathrm{Var}(N(Y))$, 利用固定 $Y = t$ 之下的条件方差公式

$$\begin{aligned}\mathrm{Var}(N(Y)|Y=t) &= \mathrm{Var}(N(t)|Y=t)\\ &= \mathrm{Var}(N(t)) \qquad Y \text{ 与 } N(t) \text{ 独立}\\ &= \lambda t\end{aligned}$$

故

$$\mathrm{Var}(N(Y)|Y) = \lambda Y, E[N(Y)|Y] = \lambda Y$$

再利用命题 5.2 的条件方差公式, 得

$$\mathrm{Var}(N(Y)) = E[\lambda Y] + \mathrm{Var}(\lambda Y) = \frac{1}{2}\lambda T + \frac{\lambda^2 T^2}{12}$$

其中, 利用了 Y 的方差 $\mathrm{Var}(Y) = T^2/12$. ■

例 5p (随机个数随机变量之和的方差) 设 $X_1, X_2, \cdots$ 是一个独立同分布的随机变量序列, N 是一取自然数的随机变量, 并且独立于序列 $\{X_i, i \geqslant 1\}$, 为计算 $\mathrm{Var}\left(\sum_{i=1}^N X_i\right)$, 先固定 N 的值作为条件

$$E\left[\sum_{i=1}^N X_i|N\right] = NE[X], \mathrm{Var}\left(\sum_{i=1}^N X_i|N\right) = N\mathrm{Var}(X)$$

其中 X 的分布与 X_i 的分布相同. 再利用条件方差公式可得

$$\mathrm{Var}\left(\sum_{i=1}^N X_i\right) = E[N]\mathrm{Var}(X) + (E[X])^2\mathrm{Var}(N)$$ ■

7.6 条件期望及预测

有时候, 在实际问题中会遇到这种情况, 某人观察到随机变量 X 的值, 基于 X 的观察值, 要对第二个随机变量 Y 的值进行预测, 通常用 $g(X)$ 表示预测值, 即当 X 的值 x 被观察到以后, $g(x)$ 就是 Y 的值的预测值. 当然, 我们希望选择 $g(X)$ 接近 Y, 选择 $g(X)$ 的一个准则是极小化 $E[(Y-g(X))^2]$. 现在我们指出在这个准则之下, Y 的最好的预测值为 $g(X) = E[Y|X]$.

命题 6.1

$$E[(Y-g(X))^2] \geqslant E[(Y-E[Y|X])^2]$$

证明:

$$\begin{aligned}E[(Y-g(X))^2|X] =& E[(Y-E[Y|X]+E[Y|X]-g(X))^2|X]\\ =& E[(Y-E[Y|X])^2|X] + E[(E[Y|X]-g(X))^2|X]\\ &+ 2E[(Y-E[Y|X])(E[Y|X]-g(X))|X]\end{aligned} \tag{6.1}$$

然而, 对于给定的 X 值, $E[Y|X]-g(X)$ 就是一个常数, 因此

$$
\begin{aligned}
&E[(Y-E[Y|X])(E[Y|X]-g(X))|X] \\
=&(E[Y|X]-g(X))E[Y-E[Y|X]|X] \\
=&(E[Y|X]-g(X))(E[Y|X]-E[Y|X])=0
\end{aligned}
\tag{6.2}
$$

这样, 由 (6.1) 式和 (6.2) 式可得

$$E[(Y-g(X))^2|X] \geqslant E[(Y-E[Y|X])^2|X]$$

上式两边再求期望即可得到命题的结论. □

注释 此处可以给出一个更加直观的证明, 当然, 在证明的严格性上要差一点. 很容易证明 $E[(Y-c)^2]$ 在 $c=E[Y]$ 时达到极小值 (见理论习题 1). 因此在我们没有任何数据可用时, 在均方误差最小的意义下, Y 的最优预测就是 $E[Y]$. 现在设得到了 X 的观察值 x, 此时预测问题与没有数据时的预测问题完全一样, 只是原来 Y 的期望改为事件 $\{X=x\}$ 之下的条件期望. 因此, Y 的最优预测是 Y 在 $X=x$ 之下的条件期望. ■

例 6a 设身高为 x 英寸的父亲, 其儿子的身高具有正态分布, 均值为 $x+1$, 方差为 4. 现设父亲的身高为 6 英尺, 试预测其儿子成长以后的身高.

解: 设父亲身高为 X, 儿子身高为 Y, 两者关系可表示为

$$Y=X+1+e$$

其中 e 为正态随机变量, 独立于 X, 并且期望为 0, 方差为 4. 对于 6 英尺的父亲, 其儿子身高的最优预测为 $E[Y|X=72]$.

$$
\begin{aligned}
E[Y|X=72]&=E[X+1+e|X=72]=73+E[e|X=72] \\
&=73+E(e) \qquad \text{利用 } X \text{ 与 } e \text{ 的独立性} \\
&=73
\end{aligned}
$$

■

例 6b 设 A 处发射一个强度为 S 的信号, 在 B 处会接收到一个强度为 R 的信号, R 是一个正态随机变量, 其期望为 S, 方差为 1. 现在设发射端的信号服从正态分布, 其期望为 μ, 方差为 σ^2. 当接收端收到的 R 的值为 r 时, 求发送信号的最优估计?

解: 首先计算发射端发送信号的 S 在给定 R 之下的条件密度

$$f_{S|R}(s|r)=\frac{f_{S,R}(s,r)}{f_R(r)}=\frac{f_S(s)f_{R|S}(r|s)}{f_R(r)}=K\mathrm{e}^{-(s-\mu)^2/2\sigma^2}\mathrm{e}^{-(r-s)^2/2}$$

其中 K 不依赖于 s. 注意下面的恒等式

$$
\begin{aligned}
&\frac{(s-\mu)^2}{2\sigma^2}+\frac{(r-s)^2}{2}=s^2\left(\frac{1}{2\sigma^2}+\frac{1}{2}\right)-\left(\frac{\mu}{\sigma^2}+r\right)s+C_1 \\
=&\frac{1+\sigma^2}{2\sigma^2}\left[s^2-2\left(\frac{\mu+r\sigma^2}{1+\sigma^2}\right)s\right]+C_1=\frac{1+\sigma^2}{2\sigma^2}\left(s-\frac{(\mu+r\sigma^2)}{1+\sigma^2}\right)^2+C_2
\end{aligned}
$$

其中 C_1, C_2 均不依赖于 s, 因此条件密度为

$$f_{S|R}(s|r) = C\exp\left\{\frac{-\left(s-\dfrac{\mu+r\sigma^2}{1+\sigma^2}\right)^2}{2\left(\dfrac{\sigma^2}{1+\sigma^2}\right)}\right\}$$

其中 C 与 s 无关. 由上式可知, 在给定 $R=r$ 之下, S 的条件分布为正态分布, 其期望和方差分别为

$$E[S|R=r]=\frac{\mu+r\sigma^2}{1+\sigma^2}, \mathrm{Var}(S|R=r)=\frac{\sigma^2}{1+\sigma^2}$$

再利用命题 6.1, 在给定 $R=r$ 之下, 对 S 的均方误差最小的估计为

$$E[S|R=r]=\frac{1}{1+\sigma^2}\mu+\frac{\sigma^2}{1+\sigma^2}r$$

由上式看出, 条件期望提供了关于 S 的信息. 它是 μ 和 r 的加权平均. 而两个权值之比为 1 与 σ^2. 其中 1 代表信号 S 发出后接收到的信号的条件方差. 而 σ^2 表示发送信号的方差. 发送信号的方差越大, μ 的作用越小 (μ 代表先验信息), 同时, 接收到信号的条件方差代表传播信号的误差 (现在为 1), 这个误差越小, 则 r 的作用越大. ■

例 6c 在数字信号处理过程中必须把连续数据离散化. 其过程如下: 取一组递增数列, $a_i, i=0,\pm1,\pm2,\cdots$, 使得 $\lim\limits_{i\to+\infty}a_i=\infty$, $\lim\limits_{i\to-\infty}a_i=-\infty$. 当 $X\in(a_i,a_{i+1}]$ 时, 选一个代表值 y_i , 这样将连续信号离散化. 用 Y 表示离散化后的值, Y 与 X 之间有如下的关系:

$$Y=y_i \qquad a_i<X\leqslant a_{i+1}$$

Y 的分布由下式给出

$$P\{Y=y_i\}=F_X(a_{i+1})-F_X(a_i)$$

现在我们的目标是要选择各区间的代表值 y_i, $i=0,\pm1,\pm2,\cdots$, 使得 $E[(X-Y)^2]$ 达到极小.

(a) 找到最优值 $y_i, i=0,\pm1,\cdots$.

对于最优的 Y, 指出

(b) $E[Y]=E[X]$, 也即均方误差最小意义下的离散化保持均值不变.

(c) $\mathrm{Var}(Y)=\mathrm{Var}(X)-E[(X-Y)^2]$.

解: (a) 对于任意的离散化随机变量 Y 利用命题 5.1 的结果, 有

$$E[(X-Y)^2]=\sum_i E[(X-y_i)^2|a_i<X\leqslant a_{i+1}]P\{a_i<X\leqslant a_{i+1}\}$$

令

$$I=i \qquad a_i<X\leqslant a_{i+1} \qquad i=0,\pm1,\cdots$$

则

$$E[(X-y_i)^2|a_i < X \leqslant a_{i+1}] = E[(X-y_i)^2|I=i]$$

利用命题 6.1 的结论, 当

$$y_i = E[X|I=i] = E[X|a_i < X \leqslant a_{i+1}] = \int_{a_i}^{a_{i+1}} \frac{xf_X(x)\mathrm{d}x}{F_X(a_{i+1}) - F_X(a_i)}$$

时, $E[(X-y_i)^2|a_i < X \leqslant a_{i+1}]$ 达到极小值. 因此, $Y = E[X|I]$ 是最优的离散化随机变量. 在最优的选择之下

(b) $E[Y] = E[E[X|I]] = E[X]$

(c) $\mathrm{Var}(X) = E[\mathrm{Var}(X|I)] + \mathrm{Var}(E[X|I])$
$= E[E[(X-Y)^2|I]] + \mathrm{Var}(Y) = E[(X-Y)^2] + \mathrm{Var}(Y)$

显然成立. ■

在某些情况下, X 和 Y 的联合分布不完全知道, 或者, 即使联合分布知道, $E[Y|X=x]$ 的计算也十分复杂. 然而, 如果我们知道 X 和 Y 的期望、方差和相关系数, 我们至少可以求出依赖于 X 的最优线性预测.

为求得 Y 的最优线性预测, 我们需要选择线性预测 $a+bX$ 的系数 a 和 b, 使得 $E[(Y-(a+bX))^2]$ 达到极小值. 为此, 先将 $E[(Y-(a+bX))^2]$ 展成一个 a, b 的多项式:

$$\begin{aligned} E[(Y-(a+bX))^2] &= E[Y^2 - 2aY - 2bXY + a^2 + 2abX + b^2X^2] \\ &= E[Y^2] - 2aE[Y] - 2bE[XY] + a^2 + 2abE[X] + b^2E[X^2] \end{aligned}$$

将上式对 a 和 b 求偏导数得到

$$\begin{aligned} \frac{\partial}{\partial a}E[(Y-(a+bX))^2] &= -2E[Y] + 2a + 2bE[X] \\ \frac{\partial}{\partial b}E[(Y-(a+bX))^2] &= -2E[XY] + 2aE[X] + 2bE[X^2] \end{aligned} \tag{6.3}$$

令偏导数为 0, 求解关于 (a, b) 的方程组 (6.3), 得到

$$\begin{aligned} b &= \frac{E[XY] - E[X]E[Y]}{E[X^2] - (E[X])^2} = \frac{\mathrm{Cov}(X,Y)}{\sigma_x^2} = \rho\frac{\sigma_y}{\sigma_x} \\ a &= E[Y] - bE[X] = E[Y] - \frac{\rho\sigma_y E[X]}{\sigma_x} \end{aligned} \tag{6.4}$$

其中 ρ 为 X, Y 的相关系数, $\sigma_y^2 = \mathrm{Var}(Y), \sigma_x^2 = \mathrm{Var}(X)$. 容易验证由 (6.4) 给出的 a, b 值使得 $E[(Y-(a+bX))^2]$ 达到极小. 因此, 在均方误差意义下, Y 的最优线性预测为

$$\mu_y + \frac{\rho\sigma_y}{\sigma_x}(X - \mu_x)$$

其中 $\mu_y = E[Y], \mu_x = E[X]$.

这个线性预测的均方误差为

$$
\begin{aligned}
&E\left[\left(Y-\mu_y-\rho\frac{\sigma_y}{\sigma_x}(X-\mu_x)\right)^2\right] \\
&=E[(Y-\mu_y)^2]+\rho^2\frac{\sigma_y^2}{\sigma_x^2}E[(X-\mu_x)^2]-2\rho\frac{\sigma_y}{\sigma_x}E[(Y-\mu_y)(X-\mu_x)] \\
&=\sigma_y^2+\rho^2\sigma_y^2-2\rho^2\sigma_y^2=\sigma_y^2(1-\rho^2) \qquad (6.5)
\end{aligned}
$$

由 (6.5) 看出, 当 ρ 接近于 $+1$ 或 -1 时, 其最优线性预测的均方误差接近于 0.

例 6d 当 X, Y 的联合分布为二元正态分布时, 由于 X 给定之下 Y 的条件期望为 X 的线性函数, 因此 Y 的最优线性预测就是最优预测, 在第 6 章例 5c 已经给出, 在正态情况

$$
E[Y|X=x]=\mu_y+\rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)
$$

■

7.7 矩母函数

随机变量 X 的矩母函数 $M(t)$ 由下式定义

$$
\begin{aligned}
M(t)&=E[\mathrm{e}^{tX}] \\
&=\begin{cases}\displaystyle\sum_x \mathrm{e}^{tx}p(x) & X\text{离散}, p(x)\text{为其分布列} \\ \displaystyle\int_{-\infty}^{\infty}\mathrm{e}^{tx}f(x)\mathrm{d}x & X\text{连续}, f(x)\text{为其密度函数}\end{cases}
\end{aligned}
$$

其中 t 为实数. 我们之所以称 $M(t)$ 为矩母函数, 其原因是 X 的所有阶矩都可以从 $M(t)$ 在 $t=0$ 的各阶微商得到. 例如

$$
M'(t)=\frac{\mathrm{d}}{\mathrm{d}t}E[\mathrm{e}^{tX}]=E\left[\frac{\mathrm{d}}{\mathrm{d}t}(\mathrm{e}^{tX})\right]=E[X\mathrm{e}^{tX}] \qquad (7.1)
$$

其中我们假定微商和期望两个运算可以交换次序, 即我们假定在离散情形下, 下式成立

$$
\frac{\mathrm{d}}{\mathrm{d}t}\left[\sum_x \mathrm{e}^{tx}p(x)\right]=\sum_x\frac{\mathrm{d}}{\mathrm{d}t}[\mathrm{e}^{tx}p(x)]
$$

在连续情形下,

$$
\frac{\mathrm{d}}{\mathrm{d}t}\left[\int \mathrm{e}^{tx}f(x)\mathrm{d}x\right]=\int\frac{\mathrm{d}}{\mathrm{d}t}[\mathrm{e}^{tx}f(x)]\mathrm{d}x
$$

这个假定在通常情况下能够验证. 特别是本书中提到的分布, 都能满足上述要求. 因此, 在 (7.1) 中, 令 $t=0$, 得

$$
M'(0)=E[X]
$$

类似地,

$$M''(t) = \frac{\mathrm{d}}{\mathrm{d}t}M'(t) = \frac{\mathrm{d}}{\mathrm{d}t}E[X\mathrm{e}^{tX}] = E\left[\frac{\mathrm{d}}{\mathrm{d}t}\left(X\mathrm{e}^{tX}\right)\right] = E[X^2\mathrm{e}^{tX}]$$

由此得到

$$M''(0) = E[X^2]$$

一般地, 对 $M(t)$ 求 n 次导数可得

$$M^{(n)}(t) = E[X^n\mathrm{e}^{tX}] \qquad n \geqslant 1$$

从而

$$M^{(n)}(0) = E[X^n] \qquad n \geqslant 1$$

现在我们对某些常见分布计算 $M(t)$.

例 7a (二项分布(参数为 (n,p))) 设 X 具有二项分布, 参数为 (n,p), 则

$$\begin{aligned} M(t) = E[\mathrm{e}^{tX}] &= \sum_{k=0}^{n} \mathrm{e}^{tk}\binom{n}{k}p^k(1-p)^{n-k} \\ &= \sum_{k=0}^{n}\binom{n}{k}(p\mathrm{e}^t)^k(1-p)^{n-k} = (p\mathrm{e}^t+1-p)^n \end{aligned}$$

最后一个等式利用了二项式展开定理. 两边求微商可得

$$M'(t) = n(p\mathrm{e}^t+1-p)^{n-1}p\mathrm{e}^t$$

故

$$E[X] = M'(0) = np$$

再求一次微商得

$$M''(t) = n(n-1)(p\mathrm{e}^t+1-p)^{n-2}(p\mathrm{e}^t)^2 + n(p\mathrm{e}^t+1-p)^{n-1}p\mathrm{e}^t$$

故

$$E[X^2] = M''(0) = n(n-1)p^2 + np$$

X 的方差为

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2 = n(n-1)p^2 + np - n^2p^2 = np(1-p)$$

这验证了之前所得的结果. ■

例 7b (泊松分布 (参数为 λ)) 设 X 的分布为泊松分布, 参数为 λ, 则

$$\begin{aligned} M(t) = E[\mathrm{e}^{tX}] &= \sum_{n=0}^{\infty}\frac{\mathrm{e}^{tn}\mathrm{e}^{-\lambda}\lambda^n}{n!} \\ &= \mathrm{e}^{-\lambda}\sum_{n=0}^{\infty}\frac{(\lambda\mathrm{e}^t)^n}{n!} = \mathrm{e}^{-\lambda}\mathrm{e}^{\lambda\mathrm{e}^t} = \exp\{\lambda(\mathrm{e}^t-1)\} \end{aligned}$$

求微商可得

$$M'(t) = \lambda e^t \exp\{\lambda(e^t - 1)\}$$
$$M''(t) = (\lambda e^t)^2 \exp\{\lambda(e^t - 1)\} + \lambda e^t \exp\{\lambda(e^t - 1)\}$$

由此可得

$$E[X] = M'(0) = \lambda$$
$$E[X^2] = M''(0) = \lambda^2 + \lambda$$
$$\mathrm{Var}(X) = E[X^2] - (E[X])^2 = \lambda$$

因此, 泊松随机变量的期望和方差均为 λ. ■

例 7c (指数分布(参数为 λ))

$$M(t) = E[e^{tX}] = \int_0^\infty e^{tx} \lambda e^{-\lambda x} dx$$
$$= \lambda \int_0^\infty e^{-(\lambda - t)x} dx = \frac{\lambda}{\lambda - t} \qquad 对t < \lambda$$

由上式可知, 指数分布的 $M(t)$ 只对 $t < \lambda$ 有定义. 对 $M(t)$ 求微商,

$$M'(t) = \frac{\lambda}{(\lambda - t)^2} \qquad M''(t) = \frac{2\lambda}{(\lambda - t)^3}$$

因此

$$E[X] = M'(0) = \frac{1}{\lambda} \qquad E[X^2] = M''(0) = \frac{2}{\lambda^2}$$

其方差为

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2 = \frac{1}{\lambda^2}$$ ■

例 7d (正态分布) 首先计算标准正态随机变量的矩母函数. 令 Z 为该标准正态随机变量,

$$M_Z(t) = E[e^{tZ}] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{tx} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty \exp\left\{-\frac{x^2 - 2tx}{2}\right\} dx$$
$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty \exp\left\{-\frac{(x-t)^2}{2} + \frac{t^2}{2}\right\} dx = e^{t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-(x-t)^2/2} dx = e^{t^2/2}$$

因此, 标准正态随机变量的矩母函数为 $M_Z(t) = e^{t^2/2}$. 对于一般正态随机变量, 只需作变换 (见 5.4 节)$X = \mu + \sigma Z$ 其中 μ, σ^2 是 X 的期望与方差, Z 为标准正态随机变量. 此时

$$M_X(t) = E[e^{tX}] = E[e^{t(\mu + \sigma Z)}] = E[e^{t\mu} e^{t\sigma Z}]$$
$$= e^{t\mu} E[e^{t\sigma Z}] = e^{t\mu} M_Z(t\sigma) = e^{t\mu} e^{(t\sigma)^2/2} = \exp\left\{\frac{\sigma^2 t^2}{2} + \mu t\right\}$$

求微商, 可得

$$M_X'(t) = (\mu + t\sigma^2)\exp\left\{\frac{\sigma^2 t^2}{2} + \mu t\right\}$$

$$M_X''(t) = (\mu + t\sigma^2)^2\exp\left\{\frac{\sigma^2 t^2}{2} + \mu t\right\} + \sigma^2\exp\left\{\frac{\sigma^2 t^2}{2} + \mu t\right\}$$

由此可得

$$E[X] = M'(0) = \mu$$

$$E[X^2] = M''(0) = \mu^2 + \sigma^2$$

因此,

$$\text{Var}(X) = E[X^2] - (E[X])^2 = \sigma^2$$ ■

表 7.1 和表 7.2 给出了某些常用的离散和连续分布的矩母函数.

表 7.1 离散概率分布

	分布列	矩母函数	均值	方差
二项分布 参数 $(n,p), 0<p<1$	$\binom{n}{x}p^x(1-p)^{n-x}$ $x=0,1,\cdots,n$	$(pe^t+1-p)^n$	np	$np(1-p)$
泊松分布 参数 $\lambda>0$	$e^{-\lambda}\dfrac{\lambda^x}{x!}$ $x=0,1,\cdots$	$\exp\{\lambda(e^t-1)\}$	λ	λ
几何分布 参数 $0<p<1$	$p(1-p)^{x-1}$ $x=1,2,\cdots$	$\dfrac{pe^t}{1-(1-p)e^t}$	$\dfrac{1}{p}$	$\dfrac{1-p}{p^2}$
负二项分布 参数 $r,p;0<p<1$	$\binom{n-1}{r-1}p^r(1-p)^{n-r}$ $n=r,r+1,\cdots$	$\left[\dfrac{pe^t}{1-(1-p)e^t}\right]^r$	$\dfrac{r}{p}$	$\dfrac{r(1-p)}{p^2}$

表 7.2 连续概率分布

	密度函数	矩母函数	均值	方差
(a,b) 上均匀分布	$f(x)=\begin{cases}\dfrac{1}{b-a} & a<x<b \\ 0 & 其他\end{cases}$	$\dfrac{e^{tb}-e^{ta}}{t(b-a)}$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$
指数分布 参数 $\lambda>0$	$f(x)=\begin{cases}\lambda e^{-\lambda x} & x\geqslant 0 \\ 0 & x<0\end{cases}$	$\dfrac{\lambda}{\lambda-t}$	$\dfrac{1}{\lambda}$	$\dfrac{1}{\lambda^2}$
Γ分布 参数 $(s,\lambda),\lambda>0$	$f(x)=\begin{cases}\dfrac{\lambda e^{-\lambda x}(\lambda x)^{s-1}}{\Gamma(s)} & x\geqslant 0 \\ 0 & x<0\end{cases}$	$\left(\dfrac{\lambda}{\lambda-t}\right)^s$	$\dfrac{s}{\lambda}$	$\dfrac{s}{\lambda^2}$
正态分布 参数 (μ,σ^2)	$f(x)=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/2\sigma^2}$ $-\infty<x<\infty$	$\exp\{\mu t+\sigma^2t^2/2\}$	μ	σ^2

矩母函数的一个重要性质是, 独立随机变量和的矩母函数等于诸随机变量各自矩母函数的乘积. 设 X 和 Y 为相互独立随机变量, 其矩母函数分别为 $M_X(t)$ 和

$M_Y(t)$. 记 $M_{X+Y}(t)$ 为 $X+Y$ 的矩母函数, 则

$$M_{X+Y}(t) = E[\mathrm{e}^{t(X+Y)}] = E[\mathrm{e}^{tX}\mathrm{e}^{tY}] = E[\mathrm{e}^{tX}]E[\mathrm{e}^{tY}] = M_X(t)M_Y(t)$$

其中倒数第二个等式利用了命题 4.1 关于独立随机变量乘积的期望的计算公式.

矩母函数的另一个重要特性是, 矩母函数唯一地确定了分布. 设 $M_X(t)$ 是 X 的矩母函数, 并在 $t=0$ 的某一邻域内有定义且有限, 则 X 的分布被 $M_X(t)$ 所唯一确定. 例如, 若

$$M_X(t) = \left(\frac{1}{2}\right)^{10}(\mathrm{e}^t+1)^{10};$$

则由表 7.1 可知 X 的分布为二项分布, 参数为 (10 , 1/2).

例 7e 设随机变量 X 的矩母函数 $M(t) = \mathrm{e}^{3(\mathrm{e}^t-1)}$, 求 $P\{X=0\}$.

解: 由表 7.1 知, $M(t) = \mathrm{e}^{3(\mathrm{e}^t-1)}$ 是泊松随机变量, 参数为 $\lambda=3$, 这样, $P\{X=0\} = \mathrm{e}^{-3}$. ■

例 7f (独立的二项随机变量的和) 设 X 和 Y 为相互独立的二项随机变量, 其分布参数分别为 (n,p) 和 (m,p), $X+Y$ 的分布是什么?

解: $X+Y$ 的矩母函数为

$$M_{X+Y}(t) = M_X(t)M_Y(t) = (p\mathrm{e}^t+1-p)^n(p\mathrm{e}^t+1-p)^m = (p\mathrm{e}^t+1-p)^{n+m}$$

然而, $(p\mathrm{e}^t+1-p)^{n+m}$ 是二项分布的矩母函数, 其相应的参数为 $(n+m,p)$. 由矩母函数唯一确定分布, 知 $X+Y$ 的分布为二项分布, 参数为 $(n+m,p)$. ■

例 7g (独立的泊松随机变量的和) 设 X,Y 为相互独立的泊松随机变量, 相应参数为 λ_1,λ_2, 求 $X+Y$ 的分布.

解: 由

$$\begin{aligned} M_{X+Y}(t) &= M_X(t)M_Y(t) \\ &= \exp\{\lambda_1(\mathrm{e}^t-1)\}\exp\{\lambda_2(\mathrm{e}^t-1)\} \\ &= \exp\{(\lambda_1+\lambda_2)(\mathrm{e}^t-1)\} \end{aligned}$$

知 $X+Y$ 的分布也是泊松分布, 分布参数为 $\lambda_1+\lambda_2$, 这验证了第 6 章例 3e 的结果. ■

例 7h (独立正态随机变量之和) 设 X 和 Y 为相互独立正态随机变量, 其参数分别为 (μ_1,σ_1^2) 和 (μ_2,σ_2^2). 则 $X+Y$ 也是正态分布, 期望为 $\mu_1+\mu_2$, 方差为 $\sigma_1^2+\sigma_2^2$.

解:

$$\begin{aligned} M_{X+Y}(t) &= M_X(t)M_Y(t) \\ &= \exp\left\{\frac{\sigma_1^2t^2}{2}+\mu_1t\right\}\exp\left\{\frac{\sigma_2^2t^2}{2}+\mu_2t\right\} \\ &= \exp\left\{\frac{(\sigma_1^2+\sigma_2^2)t^2}{2}+(\mu_1+\mu_2)t\right\} \end{aligned}$$

这个函数是期望为 $\mu_1+\mu_2$，方差为 $\sigma_1^2+\sigma_2^2$ 的正态随机变量的矩母函数. 由于矩母函数完全确定其分布函数, 故 $X+Y$ 的分布为正态分布, 其期望为 $\mu_1+\mu_2$, 方差为 $\sigma_1^2+\sigma_2^2$. ■

例 7i 计算自由度为 n 的 χ^2 分布的矩母函数.

解: 我们将具有 χ^2 分布的随机变量分解成 $Z_1^2+\cdots+Z_n^2$, 其中 $Z_1,\cdots,Z_n$ 为相互独立具有标准正态分布的随机变量, 令 $M(t)$ 为其矩母函数. 由前面论述可知

$$M(t)=(E[\mathrm{e}^{tZ^2}])^n$$

其中 Z 为标准正态随机变量. 现在,

$$\begin{aligned}E[\mathrm{e}^{tZ^2}]&=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\mathrm{e}^{tx^2}\mathrm{e}^{-x^2/2}\mathrm{d}x\\&=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\mathrm{e}^{-x^2/2\sigma^2}\mathrm{d}x \qquad 其中\ \sigma^2=(1-2t)^{-1}\\&=\sigma=(1-2t)^{-1/2}\end{aligned}$$

上述最后第二个等式利用了“密度函数之积分为 1”的结论. 因此

$$M(t)=(1-2t)^{-n/2}$$ ■

例 7j (随机个数随机变量之和的矩母函数) 设 $X_1,X_2,\cdots$ 为一独立同分布随机变量序列. 又设 N 为取值于正整数集合的随机变量, 且与 $X_i, i\geqslant 1$ 相互独立. 现在需要计算

$$Y=\sum_{i=1}^{N}X_i$$

的矩母函数. (在例 5d 中 Y 可以解释为某一天百货公司的营业额, 它是某一天顾客消费的总和, 此处每一个顾客的消费额, 以及顾客人数都是随机变量.)

首先, 求出 N 固定之下的条件期望.

$$\begin{aligned}E\left[\exp\left\{t\sum_{i=1}^{N}X_i\right\}\middle|N=n\right]&=E\left[\exp\left\{t\sum_{i=1}^{n}X_i\right\}\middle|N=n\right]\\&=E\left[\exp\left\{t\sum_{i=1}^{n}X_i\right\}\right]=[M_X(t)]^n\end{aligned}$$

其中

$$M_X(t)=E[\mathrm{e}^{tX_i}]$$

因此

$$E[\mathrm{e}^{tY}|N]=(M_X(t))^N$$

这样

$$M_Y(t)=E[(M_X(t))^N]$$

利用微分法

$$M'_Y(t) = E[N(M_X(t))^{N-1}M'_X(t)]$$

故

$$\begin{aligned} E[Y] = M'_Y(0) &= E[N(M_X(0))^{N-1}M'_X(0)] \\ &= E[NE[X]] = E[N]E[X] \end{aligned} \tag{7.2}$$

验证了例 5d 的结论. (上述论证中, 我们利用了 $M_X(0) = E[e^{0X}] = 1$ 这一结论.)

同样, 由

$$M''_Y(t) = E[N(N-1)(M_X(t))^{N-2}(M'_X(t))^2 + N(M_X(t))^{N-1}M''_X(t)]$$

得

$$\begin{aligned} E[Y^2] = M''_Y(0) &= E[N(N-1)(E[X])^2 + NE[X^2]] \\ &= (E[X])^2(E[N^2] - E[N]) + E[N]E[X^2] \qquad (7.3) \\ &= E[N](E[X^2] - (E[X])^2) + (E[X])^2E[N^2] \\ &= E[N]\text{Var}(X) + (E[X])^2E[N^2] \end{aligned}$$

由 (7.2) 和 (7.3), 得

$$\begin{aligned} \text{Var}(Y) &= E[N]\text{Var}(X) + (E[X])^2(E[N^2] - (E[N])^2) \\ &= E[N]\text{Var}(X) + (E[X])^2\text{Var}(N) \end{aligned}$$

■

例 7k 令 Y 为具有 $(0,1)$ 上均匀分布的随机变量, 假定当给定 $Y = p$ 的条件下, X 的分布为二项分布, 其参数为 (n,p). 在例 5k 中我们已经推出 X 的分布为有限点集 $\{0,1,\cdots,n\}$ 上的均匀分布. 现利用矩母函数方法建立这个结论.

解: 首先将 Y 的值固定, 在 Y 值固定的条件之下, 利用二项分布的矩母函数得到

$$E[e^{tX}|Y = p] = (pe^t + 1 - p)^n$$

由于 Y 又是 $(0,1)$ 上均匀随机变量, 对上式求期望得

$$\begin{aligned} E[e^{tX}] &= \int_0^1 (pe^t + 1 - p)^n dp = \frac{1}{e^t - 1}\int_1^{e^t} y^n dy \quad \text{作变量替换 } y = pe^t + 1 - p \\ &= \frac{1}{n+1}\frac{e^{t(n+1)-1}}{e^t - 1} = \frac{1}{n+1}(1 + e^t + e^{2t} + \cdots + e^{nt}) \end{aligned}$$

上述函数是有限集合 $\{0,1,\cdots,n\}$ 上均匀分布的随机变量的矩母函数. 由于矩母函数唯一确定其分布, 因此 X 的分布就是 $\{0,1,\cdots,n\}$ 上均匀分布. ■

联合矩母函数

我们可以把矩母函数推广到两个或多个随机变量的联合矩母函数. 设 $X_1,\cdots,X_n$ 为随机变量序列, 其联合矩母函数由下式定义:

$$M(t_1,\cdots,t_n)=E[\mathrm{e}^{t_1X_1+\cdots+t_nX_n}]$$

X_i 的矩母函数可以从联合矩母函数 $M(t_1,\cdots,t_n)$ 中得到, 即:

$$M_{X_i}(t)=E[\mathrm{e}^{tX_i}]=M(0,\cdots,0,t,0,\cdots,0)$$

其中 t 的位置刚好在第 i 个变量的地方.

可以证明, $X_1,\cdots,X_n$ 的联合矩母函数唯一地确定了它们的联合分布. 这个结论的证明已经超出了本书的范围. 利用这个结论, 可以证明, $X_1,\cdots,X_n$ 相互独立的充要条件是

$$M(t_1,\cdots,t_n)=M_{X_1}(t_1)\cdots M_{X_n}(t_n) \tag{7.4}$$

若 $X_1,\cdots,X_n$ 相互独立, 则

$$\begin{aligned}M(t_1,\cdots,t_n)&=E[\mathrm{e}^{(t_1X_1+\cdots+t_nX_n)}]=E[\mathrm{e}^{t_1X_1}\cdots\mathrm{e}^{t_nX_n}]\\&=E[\mathrm{e}^{t_1X_1}]\cdots E[\mathrm{e}^{t_nX_n}]\qquad \text{由独立性}\\&=M_{X_1}(t_1)\cdots M_{X_n}(t_n)\end{aligned}$$

另一方面, 若 (7.4) 式成立, 我们首先指出 (7.4) 式两边都是矩母函数, 等式右边与分布函数 $F_{X_1}(x_1)\cdots F_{X_n}(x_n)$ 相对应. 等式左边与 $(X_1,\cdots,X_n)$ 的分布函数 $F(X_1,\cdots,X_n)$ 相对应. 由于矩母函数唯一确定分布函数, (7.4) 说明对任意 $x_1,\cdots,x_n$, 有

$$F(x_1,\cdots,x_n)=F_{X_1}(x_1)\cdots F_{X_n}(x_n)$$

其中 F_{X_i} 为 X_i 的分布函数. 这个等式说明 $X_1,\cdots,X_n$ 是相互独立的.

例 7l　设 X 和 Y 为相互独立且同分布的正态随机变量, 期望为 μ, 方差为 σ^2, 第 6 章例 7a 证明了 $X+Y$ 与 $X-Y$ 相互独立. 现在用矩母函数的方法证明这个结论.

$$\begin{aligned}E[\mathrm{e}^{t(X+Y)+s(X-Y)}]&=E[\mathrm{e}^{(t+s)X+(t-s)Y}]=E[\mathrm{e}^{(t+s)X}]E[\mathrm{e}^{(t-s)Y}]\\&=\mathrm{e}^{\mu(t+s)+\sigma^2(t+s)^2/2}\mathrm{e}^{\mu(t-s)+\sigma^2(t-s)^2/2}\\&=\mathrm{e}^{2\mu t+\sigma^2t^2}\mathrm{e}^{\sigma^2s^2}=E[\mathrm{e}^{t(X+Y)}]E[\mathrm{e}^{s(X-Y)}]\end{aligned}$$

利用随机变量独立的充要条件 (7.4), 可知 $X+Y$ 与 $X-Y$ 是相互独立的. ■

下面我们用联合矩母函数的方法去验证第 6 章例 2b 的结论.

例 7m　假设出现的事件数是一个随机变量, 其分布为泊松分布, 参数为 λ. 又假设这些事件独立地以概率 p 被记录下来, 指出记录下来的事件数和未记录下来的事件数是相互独立的泊松随机变量, 其参数分别为 λp 与 $\lambda(1-p)$.

解: 记 X 为事件发生数, X_c 为记录下的事件数, 则未记录下来的事件数为 $X-X_c$, 首先计算条件矩母函数.

$$E[\mathrm{e}^{sX_c+t(X-X_c)}|X=n]=\mathrm{e}^{tn}E[\mathrm{e}^{(s-t)X_c}|X=n]$$
$$=\mathrm{e}^{tn}(p\mathrm{e}^{s-t}+1-p)^n=(p\mathrm{e}^s+(1-p)\mathrm{e}^t)^n$$

上述推导用到如下事实, 在 $X=n$ 的条件下, X_c 的分布为二项分布, 参数为 (n,p). 由此可得

$$E[\mathrm{e}^{sX_c+t(X-X_c)}|X]=(p\mathrm{e}^s+(1-p)\mathrm{e}^t)^X$$

上式两边求期望得

$$E[\mathrm{e}^{sX_c+t(X-X_c)}]=E[(p\mathrm{e}^s+(1-p)\mathrm{e}^t)^X]$$

由于 X 为泊松随机变量, 参数为 λ, 故 $E[\mathrm{e}^{tX}]=\mathrm{e}^{\lambda(\mathrm{e}^t-1)}$. 对任何正数 a, 可表示为 $a=\mathrm{e}^t$, 可知 $E[a^X]=\mathrm{e}^{\lambda(a-1)}$. 这样,

$$E[\mathrm{e}^{sX_c+t(X-X_c)}]=\mathrm{e}^{\lambda(p\mathrm{e}^s+(1-p)\mathrm{e}^t-1)}=\mathrm{e}^{\lambda p(\mathrm{e}^s-1)}\mathrm{e}^{\lambda(1-p)(\mathrm{e}^t-1)}$$

上式中, 令 $t=0$, 得 $E[\mathrm{e}^{sX_c}]=\mathrm{e}^{\lambda p(\mathrm{e}^s-1)}$, 类似可得 $E[\mathrm{e}^{t(X-X_c)}]=\mathrm{e}^{\lambda(1-p)(\mathrm{e}^t-1)}$, 故 X_c 与 $X-X_c$ 均为泊松随机变量, 其参数分别为 λp 和 $\lambda(1-p)$. 同时, 下式成立

$$E[\mathrm{e}^{sX_c+t(X-X_c)}]=E[\mathrm{e}^{sX_c}]E[\mathrm{e}^{t(X-X_c)}]$$

由随机变量相互独立的充要条件 (7.4) 知, X_c 与 $X-X_c$ 相互独立. ■

7.8 正态随机变量进一步的性质

7.8.1 多元正态分布

设 $Z_1,\cdots,Z_n$ 为 n 个相互独立的标准正态随机变量, 若 $X_1,\cdots,X_m$ 可以表示如下:

$$\begin{aligned}
X_1&=a_{11}Z_1+\cdots+a_{1n}Z_n+\mu_1\\
X_2&=a_{21}Z_1+\cdots+a_{2n}Z_n+\mu_2\\
&\cdots\\
X_i&=a_{i1}Z_1+\cdots+a_{in}Z_n+\mu_i\\
&\cdots\\
X_m&=a_{m1}Z_1+\cdots+a_{mn}Z_n+\mu_m
\end{aligned}$$

其中 $a_{ij},1\leqslant i\leqslant m,1\leqslant j\leqslant n$ 和 $\mu_i,1\leqslant i\leqslant m$ 均为常数. 则 $X_1,\cdots,X_m$ 称为具有多元正态分布. 利用独立正态随机变量之和仍为正态随机变量之事实可知, X_i 为正态随机变量, 其期望和方差分别为

$$E[X_i]=\mu_i,\mathrm{Var}(X_i)=\sum_{j=1}^{n}a_{ij}^2$$

现考虑它们的联合矩母函数

$$M(t_1,\cdots,t_m)=E[\exp\{t_1X_1+\cdots+t_mX_m\}]$$

由于 $\sum_{i=1}^m t_iX_i$ 本身是 $Z_1,\cdots,Z_n$ 的线性组合, 故 $\sum_{i=1}^m t_iX_i$ 为正态随机变量, 其期望和方差分别为

$$E\left[\sum_{i=1}^m t_iX_i\right]=\sum_{i=1}^m t_i\mu_i$$

$$\text{Var}\left(\sum_{i=1}^m t_iX_i\right)=\text{Cov}\left(\sum_{i=1}^m t_iX_i,\sum_{j=1}^m t_jX_j\right)=\sum_{i=1}^m\sum_{j=1}^m t_it_j\text{Cov}(X_i,X_j)$$

对于一般的正态随机变量 Y , 我们有公式

$$E[\mathrm{e}^Y]=M_Y(t)|_{t=1}=\mathrm{e}^{\mu+\sigma^2/2}$$

其中 $\mu=E[Y],\sigma^2=\text{Var}(Y)$. 将上述公式应用于 $Y=\sum_{i=1}^m t_iX_i$, 得

$$\begin{aligned}M(t_1,\cdots,t_m)&=E[\mathrm{e}^Y]=\exp\{E[Y]+\text{Var}(Y)/2\}\\&=\exp\left\{\sum_{i=1}^m t_i\mu_i+\frac{1}{2}\sum_{i=1}^m\sum_{j=1}^m t_it_j\text{Cov}(X_iX_j)\right\}\end{aligned}$$

由上式可知, $X_1,\cdots,X_m$ 的联合分布完全由 $E[X_i]$ 和 $\text{Cov}(X_i,X_j)$, $i=1,\cdots,m,j=1,\cdots,m$ 所确定.

不难指出, 当 $m=2$ 时, 此多元正态分布就是二元正态分布.

例 8a　设 X,Y 具有二元正态分布, 其参数

$$\mu_x=E[X],\mu_y=E[Y],\sigma_x^2=\text{Var}(X),\sigma_y^2=\text{Var}(Y),\rho=\text{Corr}(X,Y)$$

求 $P\{X<Y\}$.

解: 由于 $X-Y$ 也是正态分布, 其均值为

$$E[X-Y]=\mu_x-\mu_y$$

方差为

$$\begin{aligned}\text{Var}(X-Y)&=\text{Var}(X)+\text{Var}(-Y)+2\text{Cov}(X,-Y)\\&=\sigma_x^2+\sigma_y^2-2\rho\sigma_x\sigma_y\end{aligned}$$

由此可得

$$\begin{aligned}P\{X<Y\}&=P\{X-Y<0\}\\&=P\left\{\frac{X-Y-(\mu_x-\mu_y)}{\sqrt{\sigma_x^2+\sigma_y^2-2\rho\sigma_x\sigma_y}}<\frac{-(\mu_x-\mu_y)}{\sqrt{\sigma_x^2+\sigma_y^2-2\rho\sigma_x\sigma_y}}\right\}\\&=\Phi\left(\frac{\mu_y-\mu_x}{\sqrt{\sigma_x^2+\sigma_y^2-2\rho\sigma_x\sigma_y}}\right)\end{aligned}$$

■

例 8b　设 X 在给定 $\Theta=\theta$ 之下为正态随机变量, 其期望为 θ, 方差为 1. 进一

步假设 Θ 本身也是正态随机变量, 其均值为 μ, 方差为 σ^2. 求出给定 $X=x$ 之下的 Θ 的条件分布.

解: 我们不用贝叶斯公式. 我们指出 (X,Θ) 具有二元正态分布, 其联合分布密度为

$$f_{X,\Theta}(x,\theta)=f_{X|\Theta}(x|\theta)f_\Theta(\theta)$$

其中, $f_{X|\Theta}(x|\theta)$ 为正态密度函数, 期望为 θ, 方差为 1. 直接计算可以得到 $f_{X|\Theta}(x,\theta)$ 也是一个二元正态密度函数. 但是我们也可以这样看, 令 Z 为标准正态随机变量, 独立于 Θ. 则给定 $\Theta=\theta$ 之下, $Z+\Theta$ 的分布为正态分布, 期望为 θ, 方差为 1. 显然 $(Z+\Theta,\Theta)$ 与 (X,Θ) 具有相同的联合分布, 由多元正态分布的定义知, $(Z+\Theta,\Theta)$ 显然具有二元正态分布, 这样, (X,Θ) 也具有联合正态分布密度. 现在

$$E[X]=E[Z+\Theta]=\mu \quad \text{Var}(X)=\text{Var}(Z+\Theta)=1+\sigma^2$$

且

$$\begin{aligned}\rho&=\text{Corr}(X,\Theta)=\text{Corr}(Z+\Theta,\Theta)\\&=\frac{\text{Cov}(Z+\Theta,\Theta)}{\sqrt{\text{Var}(Z+\Theta)\text{Var}(\Theta)}}=\frac{\sigma}{\sqrt{1+\sigma^2}}\end{aligned}$$

由于 (X,Θ) 为二元正态分布, 因此 $X=x$ 之下 Θ 的条件分布为正态分布, 其期望为

$$E[\Theta|X=x]=E[\Theta]+\rho\sqrt{\frac{\text{Var}(\Theta)}{\text{Var}(X)}}(x-E[X])=\mu+\frac{\sigma^2}{1+\sigma^2}(x-\mu)$$

方差为

$$\text{Var}(\Theta|X=x)=\text{Var}(\Theta)(1-\rho^2)=\frac{\sigma^2}{1+\sigma^2}$$

(关于二元正态分布的条件分布的计算可见第 6 章例 5c.) ■

7.8.2 样本均值与样本方差的联合分布

设 $X_1,\cdots,X_n$ 为独立同分布正态随机变量序列, 其公共分布参数为 (μ,σ^2). 记 $\bar{X}=\sum_{i=1}^n X_i/n$ 为样本均值, 由于独立正态随机变量的和也是正态随机变量, 因此 $\bar{X}$ 也是正态随机变量, 其期望为 μ, 方差为 σ^2/n(见例 2c 和例 4a). 现在研究随机变量序列 $\bar{X},X_1-\bar{X},\cdots,X_n-\bar{X}$, 由例 4e 可知

$$\text{Cov}(\bar{X},X_i-\bar{X})=0 \qquad i=1,\cdots,n \tag{8.1}$$

现在设 Y 为正态随机变量, 且与 $X_1-\bar{X},\cdots,X_n-\bar{X}$ 相互独立, 其期望为 μ, 方差为 σ^2/n, 不难验证 $Y,X_1-\bar{X},\cdots,X_n-\bar{X}$ 为联合正态随机变量. 另一方面, 随机变量组 $\bar{X},X_1-\bar{X},\cdots,X_n-\bar{X}$ 也是联合正态的随机变量. 同时, 可以证明

$Y, X_1-\bar{X},\cdots,X_n-\bar{X}$ 的期望和协方差与 $\bar{X},X_1-\bar{X},\cdots,X_n-\bar{X}$ 的期望和协方差完全相同. 由于联合正态分布完全由其期望和协方差所确定, 因此这两组随机变量具有相同的联合分布. 这样, 就证明了 $\bar{X}$ 与 $X_1-\bar{X},\cdots,X_n-\bar{X}$ 也相互独立. 因此, $\bar{X}$ 与 $S^2=\sum_{i=1}^n(X_i-\bar{X})^2/(n-1)$ 也相互独立.

我们已经知道 $\bar{X}$ 与 S^2 相互独立, 且 $\bar{X}$ 为正态随机变量, 参数为 $(\mu,\sigma^2/n)$. 余下的, 我们只需求出 S^2 的分布. 在例 4a 中, 给出了恒等式

$$(n-1)S^2=\sum_{i=1}^n(X_i-\bar{X})^2=\sum_{i=1}^n(X_i-\mu)^2-n(\bar{X}-\mu)^2$$

两边除以 σ^2, 我们得到

$$\frac{(n-1)S^2}{\sigma^2}+\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2=\sum_{i=1}^n\left(\frac{X_i-\mu}{\sigma}\right)^2 \tag{8.2}$$

注意

$$\sum_{i=1}^n\left(\frac{X_i-\mu}{\sigma}\right)^2$$

是相互独立的标准正态随机变量的平方和, 因此其分布为自由度为 n 的 χ^2 分布, 再由例 7i, 可知其矩母函数为 $(1-2t)^{-n/2}$. 同样

$$\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2$$

的分布是自由度为 1 的 χ^2 分布, 其矩母函数为 $(1-2t)^{-1/2}$. 利用独立随机变量和的矩母函数等于各随机变量矩母函数的乘积的性质, 得

$$E[\mathrm{e}^{t(n-1)S^2/\sigma^2}](1-2t)^{-1/2}=(1-2t)^{-n/2}$$

或

$$E[\mathrm{e}^{t(n-1)S^2/\sigma^2}]=(1-2t)^{-(n-1)/2}$$

等式右边 $(1-2t)^{-(n-1)/2}$ 为 χ^2 分布的矩母函数, 自由度为 $(n-1)$, 由矩母函数唯一确定分布知, $(n-1)S^2/\sigma^2$ 的分布为自由度 $(n-1)$ 的 χ^2 分布.

综上所述, 我们有下述结论.

命题 8.1　设 $X_1,\cdots,X_n$ 为独立同分布的正态随机变量序列, 其公共分布参数为 (μ,σ^2), 则样本均值 $\bar{X}$ 与样本方差 S^2 相互独立. $\bar{X}$ 为正态随机变量, 均值为 μ, 方差为 σ^2/n; $(n-1)S^2/\sigma^2$ 为 χ^2 随机变量, 自由度为 $(n-1)$.

7.9　期望的一般定义

至此, 我们只给出了离散和连续随机变量的期望的定义. 然而, 有些随机变量

既不是离散型的, 也不是连续型的, 它们照样可以有期望值. 可以举出一个简单的例子. 设 X 为伯努利随机变量, 其分布参数为 $p=1/2$. 又设 Y 是一个 $[0,1]$ 上的具有均匀分布的随机变量, 又假定 X,Y 相互独立. 现定义一个新的随机变量

$$W=\begin{cases}X & X=1\\ Y & X\neq 1\end{cases}$$

显然, W 不是离散的 (它的取值范围为 $[0,1]$), 也不是连续的 ($P\{W=1\}=\frac{1}{2}$).

为了定义一般随机变量的期望, 我们需要给出斯蒂尔切斯积分的概念. 首先, 对于实函数 g, 积分 $\int_a^b g(x)\mathrm{d}x$ 是通过下式定义的

$$\int_a^b g(x)\mathrm{d}x=\lim\sum_{i=1}^{n}g(x_i)(x_i-x_{i-1})$$

其极限过程是这样定义的, 对于 $[a,b]$ 的分点, $a=x_0<x_1<\cdots<x_n=b$, 当 $\max\limits_{i=1,\cdots,n}(x_i-x_{i-1})\to 0$ 和 $n\to\infty$ 时, 求相应和数的极限.

对于任意一个分布函数 $F(x)$, 非负函数 $g(x)$ 在区间 $[a,b]$ 上的斯蒂尔切斯积分是这样定义的

$$\int_a^b g(x)\mathrm{d}F(x)=\lim\sum_{i=1}^{n}g(x_i)[F(x_i)-F(x_{i-1})]$$

其中的极限过程与通常积分定义中的过程一样, 即 $a=x_0<x_1<\cdots<x_n=b$, $\max\limits_{i=1,\cdots,n}(x_i-x_{i-1})\to 0$, $n\to\infty$. 进一步, 在实数轴上的斯蒂尔切斯积分由下式定义

$$\int_{-\infty}^{\infty}g(x)\mathrm{d}F(x)=\lim_{\substack{a\to-\infty\\ b\to+\infty}}\int_a^b g(x)\mathrm{d}F(x)$$

最后, 对于一般的函数 g, 定义

$$g^+(x)=\begin{cases}g(x) & g(x)\geqslant 0\\ 0 & g(x)<0\end{cases},\quad g^-(x)=\begin{cases}0 & g(x)\geqslant 0\\ -g(x) & g(x)<0\end{cases}$$

g^+ 和 g^- 都是非负函数, 称作 g 的正部和负部, $g(x)=g^+(x)-g^-(x)$. 对于 $g(x)$ 的斯蒂尔切斯积分, 可由下式定义

$$\int_{-\infty}^{\infty}g(x)\mathrm{d}F(x)=\int_{-\infty}^{\infty}g^+(x)\mathrm{d}F(x)-\int_{-\infty}^{\infty}g^-(x)\mathrm{d}F(x)$$

当 $\int_{-\infty}^{\infty}g^+(x)\mathrm{d}F(x)$ 和 $\int_{-\infty}^{\infty}g^-(x)\mathrm{d}F(x)$ 不都为 $+\infty$ 时, $\int_{-\infty}^{\infty}g(x)\mathrm{d}F(x)$ 有定义, 此时称为 $\int_{-\infty}^{\infty}g(x)\mathrm{d}F(x)$ 存在.

若 X 为任意随机变量, 其分布函数为 F, 则 X 的期望由下式定义

$$E[X]=\int_{-\infty}^{\infty}x\mathrm{d}F(x)\tag{9.1}$$

可以证明, 当 X 为离散随机变量时,

$$\int_{-\infty}^{\infty} x\mathrm{d}F(x) = \sum_{x:p(x)>0} xp(x)$$

其中 $p(x)$ 为 X 的分布列. 当 X 为连续时

$$\int_{-\infty}^{\infty} x\mathrm{d}F(x) = \int_{-\infty}^{\infty} xf(x)\mathrm{d}x$$

其中 $f(x)$ 为 X 的密度函数.

读者应该明白 (9.1) 式的直观含义. 考虑近似和

$$\sum_{i=1}^{n} x_i[F(x_i) - F(x_{i-1})]$$

将 X 的近似值 x_i 乘以 X 落入区间 $(x_{i-1}, x_i]$ 的概率, 再将这些乘积加起来, 就是 X 的期望的近似值. 当这些分割区间的长度越来越小时, 就得到 X 的期望值.

利用斯蒂尔切斯积分可以将期望的定义变得简洁, 它抓住了期望这个概念的本质. 例如, 斯蒂尔切斯积分可以将离散和连续的两种情形统一起来. 在教材中也不必分离散和连续两种情形给出定理的证明. 斯蒂尔切斯积分的性质也与通常积分性质相同. 本章中所有性质都能推广到一般情况.

小 结

设 X, Y 具有联合分布列 $p(x, y)$, 则

$$E[g(X,Y)] = \sum_{y}\sum_{x} g(x,y)p(x,y)$$

若它们具有联合密度 $f(x, y)$, 则

$$E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)\,\mathrm{d}x\,\mathrm{d}y$$

若令 $g(x, y) = x + y$, 可得

$$E[X+Y] = E[X] + E[Y]$$

这个公式可推广到

$$E\Big[\sum_{i=1}^{n} X_i\Big] = \sum_{i=1}^{n} E[X_i]$$

X, Y 的协方差由下式定义:

$$\mathrm{Cov}(X,Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$$

下面的恒等式十分有用:

$$\mathrm{Cov}\Big(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\Big) = \sum_{i=1}^{n}\sum_{j=1}^{m} \mathrm{Cov}(X_i, Y_j)$$

当 $n=m, X_i=Y_i, i=1,2,\cdots,n$ 时,

$$\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right)=\sum_{i=1}^{n}\operatorname{Var}(X_i)+2\sum_{i<j}\sum\operatorname{Cov}(X_i,X_j)$$

X,Y 之相关系数 $\rho(X,Y)$ 由下式定义:

$$\rho(X,Y)=\frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}}$$

若 X,Y 为二元离散随机变量, 则在 $Y=y$ 的条件下, X 的条件期望由下式给出:

$$E[X|Y=y]=\sum_x xP\{X=x|Y=y\}$$

如果是二元连续的, 则

$$E[X|Y=y]=\int_{-\infty}^{\infty} x f_{X|Y}(x|y)\mathrm{d}x$$

其中

$$f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}$$

为 $Y=y$ 条件下 X 的条件密度. 条件期望的性质与通常期望的性质是类似的, 只是在计算中, 所有概率都是 $Y=y$ 之下的条件概率.

记 $E[X|Y]$ 表示 Y 的函数, 当 $Y=y$ 时, 其值为 $E[X|Y=y]$. 关于条件期望的一个重要的恒等式为

$$E[X]=E[E[X|Y]]$$

在离散情形下, 是

$$E[X]=\sum_y E[X|Y=y]P\{Y=y\}$$

连续情形下, 是

$$E[X]=\int_{-\infty}^{\infty} E[X|Y=y]f_Y(y)\mathrm{d}y$$

上述的公式可以用来计算 $E[X]$. 其方法是先固定 $Y=y$, 求出条件期望 $E[X|Y=y]$, 再对 Y 求期望. 此外, 对于每一个事件 A, $P(A)=E[I_A]$, 其中 I_A 是事件 A 的示性函数. 我们也可以利用上述关于期望的计算公式来计算 $P(A)$.

X 在 $Y=y$ 之下的条件方差由下式定义:

$$\operatorname{Var}(X|Y=y)=E[(X-E[X|Y=y])^2|Y=y]$$

记 $\operatorname{Var}(X|Y)$ 表示 Y 的函数, 在 $Y=y$ 处, $\operatorname{Var}(X|Y)$ 的值是 $\operatorname{Var}(X|Y=y)$. 下面的公式称为*条件方差公式*:

$$\operatorname{Var}(X)=E[\operatorname{Var}(X|Y)]+\operatorname{Var}(E[X|Y])$$

设我们可观察到随机变量 X 的值, 我们希望根据 X 的观察值, 预测随机变量 Y 的值, 在这种情况下, $E[Y|X]$ 是使均方误差达最小的预测值.

随机变量 X 的矩母函数由下式定义:

$$M(t)=E[\mathrm{e}^{tX}]$$

X 的各阶矩都可以从 $M(t)$ 的各阶微商在 $t=0$ 处的值得到. 特别地,

$$E[X^n]=\frac{\mathrm{d}^n}{\mathrm{d}t^n}M(t)\Big|_{t=0} \qquad n=1,2,\cdots$$

矩母函数的两个重要性质是:

(i) 随机变量的矩母函数唯一地确定它的分布;

(ii) 独立随机变量和的矩母函数等于各随机变量的矩母函数的乘积.

这个结果可使下列结果的证明简化: 独立正态 (泊松或 Γ) 随机变量的和的分布仍然为正态 (泊松或 Γ) 分布.

若 $X_1,\cdots,X_m$ 均为有限个相互独立的标准正态随机变量的线性组合, 则称 $X_1,\cdots,X_m$ 为联合正态. 其联合分布由 $E[X_i],\mathrm{Cov}(X_i,X_j),i,j=1,\cdots,m$ 所确定.

设 $X_1,\cdots,X_n$ 为独立同分布的正态随机变量序列, 期望为 μ, 方差为 σ^2, 则其样本均值

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$$

和样本方差

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

相互独立. 样本均值 $\bar{X}$ 是正态随机变量, 期望为 μ, 方差为 σ^2/n; $(n-1)S^2/\sigma^2$ 是 χ^2 随机变量, 自由度为 $n-1$.

习 题

1. 一个玩游戏者同时掷一枚均匀骰子和扔一枚均匀硬币. 如果硬币正面朝上, 他赢得骰子出现点数的两倍, 否则赢得骰子点数的 1/2, 求他赢得的期望.

2. 一种名为 Clue 的牌戏是这样的, 一副牌中含三种牌. 第一种牌上画有不同的嫌疑者, 一共 6 张; 第二种牌上画有不同的武器, 也有 6 张; 第三种牌上画有不同的房间, 有 9 张. 每种牌内抽取一张, 游戏的目的是要猜出被抽出的 3 张牌.

(a) 一共有多少“副”可能的牌? (抽出的 3 张称为 1 副) 在一种游戏中, 当 3 张牌被选中以后, 每个游戏者又从剩下的牌中随机地抽取 3 张. 设某个游戏者抽到的 3 张牌中, 有 S 张嫌疑者, 有 W 张武器, 有 R 张房间, 又令 X 表示当该游戏者观察到自己的 3 张牌以后, 最早被抽出的那 3 张牌的可能“副”数.

(b) 找出 X 与 S,W,R 的关系. (c) 计算 $E[X]$.

3. 一个游戏由一局一局组成, 各局的结果都是相互独立的. 每局中游戏的双方以相同的概率赢或输一个单位. 记 W 表示某玩家的净赢的局数. 设某玩家的策略是这样的: 当他第一次赢一个单位以后立即停止游戏. 求 (a) $P\{W>0\}$, (b) $P\{W<0\}$, (c) $E[W]$.

4. 设 X, Y 的联合密度出现如下：

$$f_{X,Y}(x,y)=\begin{cases}1/y, & 0<y<1,\ 0<x<y\\ 0, & \text{其他}\end{cases}$$

求 (a) $E[XY]$, (b) $E[X]$, (c) $E[Y]$.

5. 县医院位于边长三英里的一个正方形的中心. 当发生事故时, 医院就派出救护车. 设医院位于 $(0,0)$, 发生事故地点为 (X,Y). 由于路网是格子形的, 由 $(0,0)$ 到 (x,y) 所需走的路程为 $|x|+|y|$. 假定事故是均匀地发生于这个三英里边长的正方形内, 求每次发生事故时, 救护车为到达出事地点所走的平均路程.

6. 某人掷骰子 10 次, 计算所得点数之和的期望值.

7. A、B 两人随机地, 独立地从 10 个不同的物件中选出 3 件, 求下列随机变量的期望.

(a) 两人同时选中的件数. (b) A 和 B 均不选中的件数. (c) A, B 两人中只有一人选中的件数.

8. 设有 N 个人参加某晚宴, 当一个人到达以后, 若他发现已到达的客人中有他的朋友, 则他坐在有朋友的一桌上, 否则, 他就新开一桌, 假设这 $\binom{N}{2}$ 对中, 每一对都以概率 p 成为朋友, 并且各对之间是否成为朋友是相互独立的. 求出当晚开桌数的期望值.

提示：记 $X_i=1$ 表示第 i 个到达者新开一桌, $X_i=0$ 表示第 i 个到达者与朋友坐在一桌. (假定桌子都是很大的, 可以同时坐 N 个人.)

9. 一共有 n 个球, 标号 $1,2,\cdots,n$. 还有 n 个坛子, 也标号 $1,2,\cdots,n$. 假设球 i 等可能地被放进坛子 $1,2,\cdots,i$ 中, $i=1,2,\cdots,n$. 求

(a) 空坛子的期望数;　(b) 每个坛子都不空的概率.

10. 进行 3 次试验, 每次都有相同的成功概率, 记 X 为 3 次试验中成功的次数. 若 $E[X]=1.8$.

(a) $P\{X=3\}$ 的最大可能值是什么?　(b) $P\{X=3\}$ 的最小可能值是什么?

对每种情况构造一个概率模型, 使得 $P\{X=3\}$ 取得指定的值.

提示：对于 (b), 令 U 为 $(0,1)$ 上的均匀随机变量, 然后, 根据 U 的值定义试验.

11. 考虑 n 次独立掷硬币试验, 设每次正面朝上的概率为 p. 我们称一个改变发生, 如果这一次出现的试验结果与上一次试验的结果不同. 例如, 当 $n=5$ 时, 试验结果为 HHTHT, 其中 H 表示正面朝上, 这个试验结果中现了 3 个改变. 求出 n 次试验中改变数的期望值.

提示：可将改变数表示为 $n-1$ 个伯努利随机变量的和.

12. n 个男生和 n 个女生随机地排成一队.

(a) 求在队伍中后面跟着一个女生的男生数的期望值.

(b) 假定这些学生是随机地坐在一张圆桌边, 求右边坐着一位女生的男生数的期望值.

13. 有 1000 张卡片, 上面写上编号 $1\sim1000$. 随机地分送给 1000 人, 计算卡片编号与人的年龄相符的卡片数的期望值.

14. 一个坛子内有 m 个黑球, 每次从坛子内拿走一个黑球并加进一个新球, 这个新球以概率 p 为黑球, 以概率 $1-p$ 为白球. 继续这样试验一直到拿光所有的黑球为止, 求期望试验

次数.

注: 这个模型对于理解艾滋病有帮助. 在人类免疫系统中有一类细胞称为 T 细胞. 一共有两种类型的 T 细胞, CD4 和 CD8. 在患病初期, 病人的细胞中的细胞比例与健康人的比例是相同的, 约为 60% 和 40% , 但是病人的细胞中的 CD4 比例一直在缩小. 现在的模型是这样的, 病毒入侵以后, 攻击 T 细胞, 但是身体补给系统并不区分损失的细胞是 CD4 还是 CD8, 它以 0.6 比 0.4 的比例补给. 当身体健康的时候, 这个补给比例是合适的, 由于 HIV 病毒只攻击 CD4 细胞, 这种补给方式就非常危险, 正如本题模型一样, 黑球终有一天被取光.

15. 在例 2h 中, 称 i 和 j 形成一个配对, 如果 i 拿了 j 的帽子, 而 j 拿了 i 的帽子, 求配对数的期望值.

16. 设 Z 为标准正态随机变量, 对固定的 x, 令

$$X = \begin{cases} Z & Z > x \\ 0 & \text{其他} \end{cases}$$

证明: $E[X] = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2}$.

17. 设有 n 张牌, 标以 $1, 2, \cdots, n$. 随机地排一次序, 现在让你顺次地对 n 张牌作 n 次猜测, 记 N 为正确猜测数.

(a) 若每次猜测时, 都没有以前猜测的信息, 则对任何猜测方案均有 $E[N] = 1$.

(b) 若在每次猜测以后, 都会告诉你被猜测的牌的号码. 什么是最好的猜测策略? 指出在此策略之下, 有

$$E[N] = \frac{1}{n} + \frac{1}{n-1} + \cdots + 1 \approx \int_1^n \frac{1}{x}\mathrm{d}x = \ln n$$

(c) 现在假定每次猜测以后都只告诉你对或错, 在这样的情况下下面的策略使 $E[N]$ 达到最大值. 凡是告诉你错, 就坚持原来的猜测, 直到猜对为止, 然后再换一个猜下一张卡片. 对于这个策略指出

$$E[N] = 1 + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{n!} \approx \mathrm{e} - 1$$

提示: 对于所有三个问题, 将 N 表示成示性随机变量 (也即, 伯努利随机变量) 的和.

18. 一共有 52 张的扑克牌的翻牌游戏, 每次翻一张, 直到全部翻完为止. 若第一张翻得“A”, 或第二张得 2, 或……, 或第 13 张得“K”, 或第 14 张得“A”, 等等, 我们就称出现一个配对, 求出现的配对数的期望. 注意在翻牌游戏中不分花色, 如在第 $(13n+1)$ 次翻牌中, 只要是 A(不管什么 A 的花色), 都算一个配对.

19. 某一地区一共有 r 种昆虫, 捕足昆虫时, 每次捕捉到的昆虫种类与前面的每次捕捉都是相互独立的. 现设每次捉昆虫时, 捉到第 i 种昆虫的概率为 $p_i, i=1,2,\cdots,r, \sum_{i=1}^r p_i = 1$.

(a) 指出第一次捉到种类为 1 的昆虫时, 所捉到的昆虫数的期望值.

(b) 指出第一次捉到种类为 1 的昆虫时, 所捉到昆虫种类的期望值.

注: 记 r 种昆虫的种类分别为 $1, \cdots, r$.

20. 一个坛子内有 n 个球, 第 i 个球的重量为 $W(i), i=1,2,\cdots,n$. 从坛子中无放回地取出球, 且每个球被取到的概率为其重量占剩下的球的总重量的比例. 例如, 若坛中的球为 $i_1,\cdots,i_r$, 则球 i_j 被抽到的概率为 $W(i_j)/\sum_{k=1}^{r}W(i_k), j=1,2,\cdots,r$. 计算在抽到 1 号球以前所抽球的个数的期望值.

21. 设有 100 人组成的集体, 计算

(a) 一年中刚好有 3 个人在同一天生日的天数的期望值;

(b) 100 人中不同生日的天数的期望值.

22. 掷一枚均匀骰子, 求出全部 6 个点数至少出现一次所需掷骰子的次数的期望值.

23. 1 号坛子含 5 个白球 6 个黑球, 2 号坛子含 8 白, 10 黑. 从 1 号坛子随机地取出两个球放入 2 号坛子, 然后再从 2 号坛子随机地取出 3 个球, 计算这 3 个球中白球数的期望值.

提示: 令 $X_i=1$, 如果 1 号坛子内的第 i 号白球是最后选出的 3 个球中的一个, $X_i=0$, 其他情况. 令 $Y_i=1$, 如果 2 号坛子内第 i 号白球是最后选出的 3 个球中的一个, $Y_i=0$, 其他情况. 则最后选出的 3 个球中白球的个数为 $\sum_{i=1}^{5}X_i+\sum_{i=1}^{8}Y_i$.

24. 一个瓶中含有两种药片, 大的 m 片, 小的 n 片. 每天, 病人随机地从中选一片, 如果选到小的, 就吃下去, 如果选到的是大的, 就掰成两半, 吃掉一半, 剩下一半成为小药片, 放进瓶内.

(a) 令 X 表示拿到最后一片大的药片并且将吃剩的半片放回瓶中以后, 瓶子内剩下的小药片的数目, 求 $E[X]$.

提示: 定义 $n+m$ 个示性随机变量, 一类是原来 n 个小药片, 另一类是 m 个大药片吃掉一半以后成为的小药片, 再利用, 例 2m 中的方法.

(b) 记 Y 表示病人拿到最后一片大药片的天数, 求 $E[Y]$.

提示: X 与 Y 之间有什么关系?

25. 设 $X_1,X_2,\cdots$ 是独立同分布的连续随机变量序列, 记 $N\geqslant 2$ 满足 $X_1\geqslant X_2\geqslant\cdots\geqslant X_{N-1}<X_N$, 即 N 是这个序列开始上升的时刻. 证明 $E[N]=\mathrm{e}$.

提示: 首先计算 $P\{N\geqslant n\}$.

26. 设 $X_1,X_2,\cdots,X_n$ 为独立同分布并在 $(0,1)$ 上均匀分布的随机变量序列, 计算

(a) $E[\max(X_1,\cdots,X_n)]$;　(b) $E[\min(X_1,\cdots,X_n)]$.

***27.** 一共有 101 个物品放入 10 个盒子, 显然至少有一个盒子包含多于 10 个物品, 用概率方法证明此结果.

***28.** 设一个由 n 个元件组成的系统, 这 n 个元件排成一个圆周. 每一个元件有两种状态, 失效和工作. 若这个 n 个元件组成的系统中, 不存在连续 r 个元件, 其中至少有 k 个失效, 则这个系统正常运行, 这种系统称为圆周的 (k,r,n) 可靠性系统, $k\leqslant r\leqslant n$. 现在共有 47 个元件, 其中 8 个失效, 指出不可能安排出一个正常运行的 $(3,12,47)$ 圆周系统.

***29.** 一共有四种不同的优惠券, 前两种优惠券组成一组, 后两种优惠券组成另一组 (例如前两种为食品优惠券, 后两种为服装优惠券.) 每得到一个新的优惠券, 它是第 i 种的概率为 $p_i, i=1,2,3,4$. 其中 $p_1=p_2=1/8, p_3=p_4=3/8$. 记 X 为达到下列目的之一所需收集的优惠券数, 并求 $E[X]$.

(a) 所有 4 种优惠券; (b) 第一组中所有的各种优惠券;

(c) 第二组中所有的各种的优惠券; (d) 第一组或第二组中所有各种的优惠券.

30. 设 X, Y 为独立同分布的随机变量, 其公共期望为 μ, 方差为 σ^2. 求 $E[(X-Y)^2]$.

31. 计算习题 6 中点数之和的方差.

32. 在习题 9 中, 计算空坛子数的方差.

33. 设 $E[X]=1, \operatorname{Var}(X)=5$, 求 (a) $E[(2+X)^2]$; (b) $\operatorname{Var}(4+3X)$.

34. 10 对夫妇随机地坐在一张圆桌上, 记 X 为夫妇坐在相邻位置的对数, 求

(a) $E[X]$. (b) $\operatorname{Var}(X)$.

35. 一副扑克牌, 一张一张翻开, 分别计算为出现下列牌所需翻牌数的期望值.

(a) 两个"A"; (b) 5 个黑桃; (c) 13 张红桃.

36. 掷一枚均匀骰子 n 次, 记 X 和 Y 分别是出现 1 点和 2 点的次数, 计算 $\operatorname{Cov}(X,Y)$.

37. 掷一枚均匀骰子两次, 记 X 为两次出现的点数之和, 记 Y 为第一次出现的点数减去第二次出现的点数的差, 计算 $\operatorname{Cov}(X,Y)$.

38. 设 X, Y 具有联合密度函数

$$f(x,y)=\begin{cases}2\mathrm{e}^{-2x}/x & 0\leqslant x<\infty, 0\leqslant y\leqslant x\\ 0 & \text{其他}\end{cases}$$

计算 $\operatorname{Cov}(X,Y)$.

39. 设 $X_1, X_2, \cdots$ 为独立同分布随机变量序列, 其公共期望为 μ, 公共方差为 σ^2. 记 $Y_n = X_n + X_{n+1} + X_{n+2}$, 求 $\operatorname{Cov}(Y_n, Y_{n+j})$, $j \geqslant 0$.

40. 设 X, Y 的联合密度为

$$f(x,y)=\frac{1}{y}\mathrm{e}^{-(y+x/y)} \qquad x>0, y>0$$

求 $E[X], E[Y]$, 并证明 $\operatorname{Cov}(X,Y)=1$.

41. 鱼池中共有 100 条鱼, 其中 30 条为鲤鱼, 现抓住 20 条, 求这 20 条中鲤鱼数的期望与方差. 计算时, 你作了什么样的假定?

42. 由 20 人组成的集体, 其中 10 人为男生, 10 人为女生. 他们随机地分成 10 组, 每组 2 人, 求 10 组中男女混合的组数的期望与方差. 若 20 人由 10 对夫妻组成, 问分组中夫妻组的组数的期望和方差.

43. 设 $X_1, \cdots, X_n$ 为独立同分布的随机变量序列, 其公共分布函数为未知的连续函数 $F(x)$. 又设 $Y_1, \cdots, Y_m$ 也是独立同分布的随机变量序列, 其公共分布函数为未知的连续函数 $G(y)$. 现将这 $n+m$ 个变量排序, 令

$$I_i=\begin{cases}1 & \text{若 } n+m \text{ 个值中第 } i \text{ 个最小的值来自} X\\ 0 & \text{其他}\end{cases}$$

$R=\sum_{i=1}^{n+m} iI_i$ 称为样本 X 的秩和, 它是检验 F 与 G 是否相同的标准统计量. (这种检验称为威尔科克森 (Wilcoxon) 秩和检验), 当 R 不是很大也不是很小时, 接受假设 $F=G$. 现在在 $F=G$ 的假定下, 计算 R 的期望和方差.

提示: 利用例 3e 的结果.

44. 有两种生产工艺生产某种产品, 由工艺 i 生产的产品的质量的分布为连续分布 $F_i, i=1,2$. 现设利用工艺 1 生产 n 件产品, 工艺 2 生产 m 件产品, 令

$$X_j = \begin{cases} 1 & \text{若第 } j \text{ 个最好产品来自于工艺 } 1 \\ 2 & \text{其他} \end{cases}$$

$X_1, \cdots, X_{n+m}$ 中有 n 个 1, m 个 2. 记 R 表示"1"的游程的个数, 例如, $n=5, m=2, X=1,2,1,1,1,1,2$, 则 $R=2$. 在 $F_1=F_2$ 的假定之下, R 的期望和方差是什么? (如果两种工艺生产的产品质量完全一致, 则 $F_1=F_2$.)

45. 设 X_1, X_2, X_3, X_4 是两两不相关的随机变量序列, 具有均值 0 和方差 1. 计算下列各对随机变量的相关系数: (a) X_1+X_2 和 X_2+X_3; (b) X_1+X_2 和 X_3+X_4.

46. 在赌场中有这样的游戏: 赌徒 1 和赌徒 2 先轮流掷两枚骰子, 然后庄家也掷两枚骰子. 若赌徒 i 的点数和严格地大于庄家的点数和, 则赌徒 i 赢, 令

$$I_i = \begin{cases} 1 & \text{若赌徒 } i \text{ 赢} \\ 0 & \text{若赌徒 } i \text{ 输} \end{cases}$$

指出 I_1, I_2 为正相关的, 解释这个结果的原因.

47. 考虑有 n 个节点 $1,2,\cdots,n$ 的一个图, 假设 $\binom{n}{2}$ 对节点中的每一对节点独立地以概率 p 连上一条边而相互联结. 对于每一个节点 i, 它所联结的点的个数 D_i 称为该点的阶.

(a) D_i 的分布是什么? (b) 计算相关系数 $\rho(D_i, D_j)$.

48. 连续地掷一枚均匀骰子, 记 X 和 Y 分别为得到 6 点和 5 点所需掷的次数, 求

(a) $E[X]$; (b) $E[X|Y=1]$; (c) $E[X|Y=5]$.

49. 在一个盒子里有两枚不均匀的硬币, 在抛掷硬币时正面朝上的概率分别为 0.4 和 0.7. 现在从中随机地取一枚硬币, 连续抛掷 10 次. 已知前 3 次抛掷硬币时, 有 2 次正面朝上, 问 10 次抛掷后, 正面朝上的次数的条件期望是多少?

50. 设 X, Y 的联合密度为

$$f(x,y) = \frac{e^{-x/y}e^{-y}}{y} \qquad 0<x<\infty, 0<y<\infty$$

计算 $E[X^2|Y=y]$.

51. X, Y 的联合密度为

$$f(x,y) = \frac{e^{-y}}{y} \qquad 0<x<y, 0<y<\infty$$

计算 $E[Y^3|Y=y]$.

52. 设一个班由 r 个不相交的小组组成, 记 p_i 为第 i 个小组占班上总人数的比例, $i=1,2,\cdots,r$. 若第 i 个小组的成员的平均体重为 $w_i, i=1,2,\cdots,r$. 问全班成员的平均体重是多少?

53. 一囚犯发现三个暗道, 从第一个道出去经过 2 天会回到原地, 从第二个暗道出去经过 4 天会回到原地, 从第三个暗道经过一天会走出监狱. 假定该犯人总是按 0.5, 0.3 和 0.2 的概率选择暗道, 问他逃出监狱所需的平均天数.

54. 考虑一个掷骰子游戏, 每次掷一对骰子, 如果点数和为 7 游戏结束, 赢得为 0. 如果点数和不是 7, 那么可作两种选择, 你可以将点数和作为赢得而停止游戏, 或者重新开始游戏. 对每一个 $i, i = 2, \cdots, 12$, 可以制定一个游戏策略, 当掷得点数和不等于 7, 但这个数小于 i 时, 选择重新开始游戏. 只有当点数和 $\geqslant i$ 时才停止游戏, 求此时的期望赢得. i 为何值时, 能获得最大的期望赢得?

提示: 记 X_i 为采用策略 i 时的赢得. 将首次掷骰子所得的点数和作为条件, 求 X_i 的条件期望.

55. 10 个猎人在等着野鸭飞过天空, 当野鸭飞过时, 猎人们随机地选中一个目标, 同时射击. 假定猎人们随机地选择目标, 且他们之间相互独立, 各自以 0.6 的概率击中目标, 又假定飞过的野鸭数的分布为泊松分布, 其参数为 $\lambda = 6$. 求被击中的鸭子数目的期望值. (泊松随机变量取 0 值的情况不予考虑.)

56. 设某大楼共有 N 层, 在 1 层有 n 个人进入电梯, n 是泊松随机变量, 其分布参数为 $\lambda = 10$. 设对于每一个人, 其目的地是相互独立的, 并且是等可能地进入每一楼层. 求该电梯在 2 层以上平均停多少次才能将乘客送到他们各自的楼层.

57. 设某工厂每周平均发生 5 次事故, 在每次事故中受伤的平均人数为 2.5. 并且假设各次事故中受伤人数是相互独立的, 计算一周内受伤的工人数的期望值.

58. 设抛掷硬币时, 正面朝上的概率为 p, 不断地抛掷硬币, 直到正面和反面都曾出现为止. 求: (a) 抛掷硬币次数的期望; (b) 最后一次出现正面朝上的概率.

59. 一共有 $n+1$ 个人参加游戏, 每个人以概率 p 取胜, 且各人的输赢是相互独立的, 所有胜者均分一个单位的奖金 (例如, 若有 4 个人赢, 则每个人得到 1/4 , 如果没有人赢, 则无人得奖). 设 A 为某参加者, X 代表 A 所得的奖励.

(a) 计算游戏者所获得的总奖励的期望值. (b) 指出 $E[X] = [1-(1-p)^{n+1}]/(n+1)$.

(c) 以 A 是否赢作为条件, 计算 $E[X]$. 并证明

$$E[(1+B)^{-1}] = \frac{1-(1-p)^{n+1}}{(n+1)p}$$

其中 B 为二项随机变量, 其参数为 (n, p).

60. $m+2$ 个人玩游戏, 每一个人拿出一元钱放在一起以便开始游戏. 连续抛掷一枚均匀硬币 n 次, 其中 n 为奇数, 并记录每次抛掷结果. 在抛掷硬币之前, 参加游戏的每一个人, 必须写下一个他对每次抛掷结果的猜测, 例如, 当 $n = 3$ 时, 游戏参加者写下 (HHT) 表示他预测第一次会出现正面朝上, 第二次正面朝上, 第三次出现反面朝上. 当真正的 n 次抛掷结果出来以后, 每个人核对 n 次抛掷结果和他的猜测结果, 计算他正确猜测的次数. 例如, 若实际抛掷结果为 (HHH) 而他的猜测为 (HHT), 此时, 他正确猜测数为 2. 作为对猜测者的奖励, 将 $m+2$ 元钱平均分配给那些正确猜测数最多的人. 由于抛掷一个硬币出现 H 或 T 的概率是相同的, 而各次抛掷又相互独立, 因此其中 m 个参加者决定随机地预测, 特别地, 他们在事先可以自己抛掷一枚硬币 n 次, 将所得的结果作为猜测. 然而, 另外两个人决定合伙, 其中一个人与前面 m 个人的做法完全一样, 另一个人预测刚好其合伙人相反的结果. 例如, 若一个人预测为 (HHT), 则另一个人的预测为 (TTH).

(a) 指出两个合伙人中至少有一个人的正确猜测数大于 $n/2$(n 为奇数).

(b) 记 X 为不参加合伙的 m 个人中正确猜测数超过 $n/2$ 的人数. X 的分布是什么?

(c) 设 X 如 (b) 中所定义, 指出

$$E[\text{合伙人所得钱数}] = (m+2)E\left[\frac{1}{X+1}\right]$$

(d) 利用习题 59(c), 指出

$$E[\text{合伙人所得钱数}] = \frac{2(m+2)}{m+1}\left[1-\left(\frac{1}{2}\right)^{m+1}\right]$$

并具体计算当 $m=1,2,3$ 时的值. 由于可以证明

$$\frac{2(m+2)}{m+1}\left[1-\left(\frac{1}{2}\right)^{m+1}\right] > 2$$

这说明合伙策略可以得到更多的期望收入.

61. 设 $X_1, X_2, \cdots$ 为独立同分布随机变量, 其公共分布为 F. 设 N 为与它们独立的随机变量, 具有几何分布, 参数为 p, 令 $M=\max(X_1,\cdots,X_N)$

(a) 求 $P\{M\leqslant x\}$(以 N 作为条件, 利用条件期望进行计算); (b) 求 $P\{M\leqslant x|N=1\}$;

(c) 求 $P\{M\leqslant x|N>1\}$; (d) 利用 (b), (c) 再导出 (a).

62. 令 $U_1, U_2, \cdots$ 为一独立同分布的随机变量序列, 其公共分布为 $(0,1)$ 上均分布. 在例 5i 中我们已经指出对于 $0\leqslant x\leqslant 1, E[N(x)]=\mathrm{e}^x$, 其中

$$N(x)=\min\left\{n:\sum_{i=1}^{n}U_i>x\right\}.$$

此处给出这一结论的另一证法

(a) 利用对 n 归纳法, 指出对 $0<x\leqslant 1$ 和一切 $n\geqslant 0$

$$P\{N(x)\geqslant n+1\}=\frac{x^n}{n!}$$

提示: 首先以 U_1 为条件, 然后利用归纳假设.

(b) 利用 (a) 证明 $E[N(x)]=\mathrm{e}^x$.

63. 一个坛子里含 30 个球, 其中 10 个红球, 8 个蓝球, 从坛子里随机地取出 12 个球, 记 X 为红球数, Y 为蓝球数, 求 $\mathrm{Cov}(X,Y)$.

(a) 定义适当的示性 (也即伯努利) 随机变量 X_i, Y_i, 使得 $X=\sum_{i=1}^{10}X_i, Y=\sum_{j=1}^{8}Y_j$.

(b) 利用条件期望的方法求 $E[XY]$. (以 $X=x$ 作为条件或以 $Y=y$ 作为条件.)

64. 设某箱子有两种灯泡, 第 i 种灯炮的寿命的均值为 μ_i, 标准差为 $\sigma_i, i=1,2$. 现从箱中随机地抽取一灯炮, 抽到第 1 种的概率为 p, 抽到第 2 种的概率为 $1-p$. 记抽出的灯泡寿命为 X, 求 (a) $E[X]$; (b) $\mathrm{Var}(X)$.

65. 已知气候良好时一年冬季风暴次数是均值为 3 的泊松随机变量, 而气候恶劣的年头冬季风暴个数的是均值为 5 的泊松随机变量. 若下一个年头气候好的概率为 0.4, 不好的概率为 0.6. 求出下一个冬季的风暴数的期望值和方差.

66. 在例 5c 中, 计算矿工到达安全地点所需时间的方差.

67. 设有一赌徒, 他在每次赌博中, 赢和输的概率分别为 p 和 $1-p$. 当 $p>1/2$ 时, 有一种 Kelly 策略, 就是总用现有财产的 $2p-1$ 倍作赌注. 试计算从他的初始财产 x 算起, 利用 Kelly 策略, 经过 n 次赌博以后的期望财产.

68. 一个人在一年中事故的次数为泊松随机变量, 参数为 λ, 但是事故的发生次数随人而变. 设人群中有 60% 的人的参数为 $\lambda=2$, 另有 40% 的人的参数为 $\lambda=3$. 现从中随机地找一人, 他在一年中 (a) 有 0 个事故 (b) 有 3 个事故的概率是多少? 若已知他上一年没有事故, 下一年有 3 个事故的概率有多大?

69. 在习题 68 中, 设参数 λ 本身为一随机变量, 人群中参数 $\lambda<x$ 的比例为 $1-\mathrm{e}^{-x}$, 重复该问题的计算.

70. 设一坛子中有很多硬币, 假定抛掷硬币时正面朝上的概率为 p, 而 p 的值与拿到的硬币有关. 由于从坛子中随机地取一硬币, p 也可以看成随机变量, 并且假定 p 的值在 $[0,1]$ 上均匀分布. 现在从坛子中随机地取一硬币, 连续抛掷两次, 计算下列事件的概率:

(a) 第一次抛掷正面朝上; (b) 两次均为正面朝上.

71. 假定在习题 70 中, 抛掷硬币 n 次. 记 X 为正面朝上出现的次数. 证明

$$P\{X=i\}=\frac{1}{n+1}\qquad i=0,1,\cdots,n$$

提示, 利用公式

$$\int_0^1 x^{a-1}(1-x)^{b-1}\mathrm{d}x=\frac{(a-1)!\,(b-1)!}{(a+b-1)!}$$

其中 a,b 为正整数.

72. 设在问题 70 中, 随机的取出一枚硬币, 并连续抛掷, 直到出现正面朝上为止. 记 N 为抛掷次数, 计算 (a) $P\{N\geqslant i\},i\geqslant 0$; (b) $P\{N=i\}$; (c) $E[N]$.

73. 在例 6b 中, 令 S 代表发送的信号, R 代表接收的信号.

(a) 计算 $E[R]$. (b) 计算 $\mathrm{Var}(X)$. (c) R 是否为正态随机变量? (d) 计算 $\mathrm{Cov}(R,S)$.

74. 在例 6c 中, 假定 X 的分布为 $(0,1)$ 上均匀分布. 并假定 $(0,1)$ 区间的离散化由 $a_0=0,a_1=1/2,a_2=1$ 所确定. 找出最优的离散化随机变量 Y, 并计算相应的 $E[(X-Y)^2]$.

75. 设 X 和 Y 的矩母函数分别为 $M_X(t)=\exp\{2\mathrm{e}^t-2\}$ 和 $M_Y(t)=(\frac{3}{4}\mathrm{e}^t+\frac{1}{4})^{10}$. 若 X 和 Y 相互独立, 计算 (a) $P\{X+Y=2\}$; (b) $P\{XY=0\}$; (c) $E[XY]$?

76. 设某人掷两次骰子, 记 X 为第一次得到的点数, Y 为两次的点数之和. 计算 X 和 Y 的联合矩母函数.

77. 设 X,Y 的联合密度为

$$f(x,y)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-y}\mathrm{e}^{-\frac{(x-y)^2}{2}}\qquad 0<y<\infty,-\infty<x<\infty$$

(a) 计算 X,Y 的联合矩母函数. (b) 分别计算 X 和 Y 的矩母函数.

78. 两个信封里各有一张支票. 你随便打开一个信封, 看到了支票上的钱数. 此时, 你可以有两种选择, 或者接受这张支票, 或者接受另一个信封内的支票, 你该怎么办? 是否可以找到一种办法, 比接受第一张支票更好?

记 A,B 为两张指标的面值, $A<B$, 若随机地选一张并接受它, 你的期望收入为 $(A+B)/2$. 考虑下面的第二种策略: 令 $F(x)$ 为严格上升并且连续的分布函数, 设第一张支票的面值为 x, 然后以 $F(x)$ 的概率接受, 以 $1-F(x)$ 的概率改变为接受另一张支票.

(a) 证明, 如果使用第二种策略, 其期望收入将大于 $(A+B)/2$.

提示: 以第一张支票的面值为 A 或 B 作为条件.

现在考虑另一个 $x-$ 策略, 其中 x 为固定的值, 若第一张支票的面值大于 x, 则接受, 否则接受另一个信封的支票.

(b) 证明, 对于任何 x 值, 在 $x-$ 策略之下, 你所得到的期望收入至少是 $(A+B)/2$, 特别地, 当 x 在 A, B 之间的时候, 你所得到的期望面值大于 $(A+B)/2$.

(c) 记 X 为 $(-\infty, \infty)$ 上的连续随机变量, 且 $P\{A < X < B\} > 0$. 考虑如下的策略: 产生一个 X 的值 x, 然后利用 (b) 中的 $x-$ 策略, 证明你所得到的期望收入大于 $(A+B)/2$.

79. 设 X, Y 为连续的两个星期的销售量 (以 1000 元为单位), 已知 (X, Y) 具有二元正态分布, 其公共期望 40, 公共标准差为 6, 相关系数为 0.6.

(a) 求出两星期总销售额超过 90 的概率.

(b) 若相关系数由 0.6 减为 0.2, 你认为 (a) 中的概率会增加还是减少? 说明你的理由.

(c) 相关系数为 0.2 时重复 (a) 的计算.

理 论 习 题

1. 证明 $E[(X-a)^2]$ 在 $a = E[X]$ 时达到最小.

2. 设 X 为连续随机变量, 密度函数为 f, 证明 $E[|X-a|]$ 在 a 等于 F 的中位数时达到最小值.

提示: 注意

$$E[|X-a|] = \int |x-a| f(x) \mathrm{d}x$$

将积分区域分为两部分 $x < a$ 和 $x > a$, 然后求微商.

3. 在下列情况下证明命题 2.1:

(a) (X, Y) 具有联合分布列;

(b) (X, Y) 具有联合概率密度并且函数 $g(x, y) \geqslant 0$ 对一切 x, y 成立.

4. 设 X 具有有限期望 μ 和方差 σ^2, 记 $g(\cdot)$ 为二次可微函数, 证明

$$E[g(X)] \approx g(\mu) + \frac{1}{2} g''(\mu) \sigma^2$$

提示: 将 $g(\cdot)$ 在 μ 处展开, 利用前三项, 忽略余项.

5. 设 $A_1, A_2, \cdots, A_n$ 为任意事件, 定义 $C_k = \{A_i\text{中至少}k\text{个事件发生}\}$, 证明

$$\sum_{k=1}^{n} P(C_k) = \sum_{k=1}^{n} P(A_k)$$

提示: 记 X 为 A_i 中发生的事件数, 指出等式两边均为 $E[X]$.

6. 注意到, 在文中有等式

$$E\Big[\sum_{i=1}^{\infty} X_i\Big] = \sum_{i=1}^{\infty} E[X_i]$$

其中 X_i 为非负随机变量, 由于积分是和的极限, 我们希望下式

$$E\left[\int_0^{\infty} X(t)\mathrm{d}t\right] = \int_0^{\infty} E[X(t)]\mathrm{d}t$$

对所有非负随机变量 $X(t), 0 \leqslant t < \infty$ 成立. 此式确实成立, 利用此结论给出

$$E[X] = \int_0^{\infty} P\{X > t\}\mathrm{d}t$$

的另一证明 (其中 X 为非负随机变量)

提示: 对一切 $t \geqslant 0$, 定义

$$X(t) = \begin{cases} 1 & t < X \\ 0 & t \geqslant X \end{cases}$$

然后利用公式 $X = \int_0^{\infty} X(t)\mathrm{d}t$.

7. X 称为随机地大于 Y(记作 $X \geqslant_{\mathrm{st}} Y$), 如果对所有 t, 有下式成立

$$P\{X > t\} \geqslant P\{Y > t\}$$

指出在下列两种情况下, 若 $X \geqslant_{\mathrm{st}} Y$, 则 $E[X] \geqslant E[Y]$.

(a) X, Y 为非负随机变量. (b) X, Y 为任意随机变量.

提示: 记

$$X = X^+ - X^-$$

其中

$$X^+ = \begin{cases} X & X \geqslant 0 \\ 0 & X < 0 \end{cases} \qquad X^- = \begin{cases} 0 & X \geqslant 0 \\ -X & X < 0 \end{cases}$$

类似地, 将 Y 写成 $Y^+ - Y^-$, 然后利用 (a).

8. 指出 $X \geqslant_{\mathrm{st}} Y$ 成立的充要条件是下式

$$E[f(X)] \geqslant E[f(Y)]$$

对所有增函数 f 成立.

提示: 由 $X \geqslant_{\mathrm{st}} Y$ 推导 $E[f(X)] \geqslant E[f(Y)]$, 只需指出 $f(X) \geqslant_{\mathrm{st}} f(Y)$ 然后利用理论习题 7, 证明逆命题只需定义适当的增函数 f.

9. 设有一硬币, 在抛掷时正面朝上的概率为 p. 现在抛掷 n 次, 计算正面朝上的游程的长度为 k 的个数的期望值, $1 \leqslant k \leqslant n$.

10. 设 $X_1, X_2, \cdots, X_n$ 为独立同分布的正随机变量, 计算

$$E\left[\left(\sum_{i=1}^{k} X_i\right) \Big/ \left(\sum_{i=1}^{n} X_i\right)\right] \qquad k \leqslant n$$

11. 考虑 n 次独立重复试验, 每次试验可有 r 个不同的结果, 其相应的概率为 $P_1, P_2, \cdots, P_r$. 记 X 表示在 n 次试验中从未出现的结果的个数, 计算 $E[X]$, 并且指出在 $P_1, P_2, \cdots, P_r$ 的所有可能取值中, 当 $P_i = 1/r, i = 1, \cdots, r$ 时, $E[X]$ 达到极小值.

12. 设 $X_1, X_2, \cdots$ 为独立同分布随机变量序列, X_n 的分布列为

$$P\{X_n = 0\} = P\{X_n = 2\} = \frac{1}{2}, \quad n \geqslant 1$$

随机变量 $X = \sum_{n=1}^{\infty} X_n/3^n$ 的分布称为**康托分布**. 求 $E[X]$ 和 $\mathrm{Var}(X)$.

13. 设 $X_1, \cdots, X_n$ 为独立同分布的连续随机变量序列, 我们称在 j 时打破记录, 若 $X_j \geqslant X_i$, 对一切 $1 \leqslant i \leqslant j$ 成立. 证明:

(a) E[打破记录数] $= \sum_{j=1}^{n} 1/j$; (b) Var(打破记录数) $= \sum_{j=1}^{n} (j-1)/j^2$.

14. 在例 2i 中, 记 X 为集齐一套优惠券所需收集的优惠券张数. 证明

$$\operatorname{Var}(X) = \sum_{i=1}^{N-1} \frac{iN}{(N-i)^2}$$

当 N 充分大时, 这个数近似于 $N^2\pi^2/6$(即 $\operatorname{Var}(X)/(N^2\pi^2/6) \to 1, N \to \infty$).

15. 设有 n 次独立试验, 第 i 次成功的概率为 P_i.

(a) 计算成功次数的期望值, 并记为 μ.

(b) 对固定的 μ 值, 选择 $P_1, \cdots, P_n$ 的值使成功次数的方差达极大.

(c) 什么样的选择使方差达极小?

***16.** 设 $S = \{1, 2, \cdots, n\}$ 为一个集合, 集合中的任一元素可以被涂成红色或蓝色. 又设 $A_1, \cdots, A_r$ 为 S 的 r 个子集. 指出, 存在一种涂颜色的方法, 使得 $A_1, \cdots, A_r$ 中具有相同颜色元素的子集数不会超过 $\sum_{i=1}^{r} (1/2)^{|A_i|-1}$, 其中 $|A_i|$ 表示集合 A_i 中元素的个数.

17. 设 X_1, X_2 相互独立, 具有公共的期望值 μ. 设 $\operatorname{Var}(X_1) = \sigma_1^2, \operatorname{Var}(X_2) = \sigma_2^2$. 由于 μ 未知, 我们希望以 X_1, X_2 的加权平均作为 μ 的估计, 即以 $\lambda X_1 + (1-\lambda)X_2$ 作为 μ 的估计. 什么样的 λ 值使估计的方差达最小? 解释所得结果的合理性.

18. 在例 4f 中, 我们在求多项随机变量 N_i 和 N_j 的协方差 $(-mP_iP_j)$ 时, 将 N_i 和 N_j 表示成示性随机变量之和. 这个结果也可从下式得到:

$$\operatorname{Var}(N_i + N_j) = \operatorname{Var}(N_i) + \operatorname{Var}(N_j) + 2\operatorname{Cov}(N_i N_j)$$

(a) $N_i + N_j$ 的分布是什么? (b) 利用上面的等式, 证明 $\operatorname{Cov}(N_i, N_j) = -mP_iP_j$.

19. 如果 X 和 Y 同分布, 但不一定独立, 证明

$$\operatorname{Cov}(X+Y, X-Y) = 0$$

20. *条件协方差公式*(conditional covariance formula). 对于给定 Z, X 和 Y 的条件协方差由下式定义

$$\operatorname{Cov}(X, Y|Z) \equiv E[(X - E[X|Z])(Y - E[Y|Z])|Z]$$

(a) 证明

$$\operatorname{Cov}(X, Y|Z) = E[XY|Z] - E[X|Z]E[Y|Z]$$

(b) 证明条件协方差公式

$$\operatorname{Cov}(X, Y) = E[\operatorname{Cov}(X, Y|Z)] + \operatorname{Cov}(E[X|Z], E[Y|Z])$$

(c) 在 (b) 中令 $X = Y$, 就可得到条件方差公式.

21. 令 $X_{(i)} (i = 1, 2, \cdots, n)$ 为 $(0, 1)$ 上均匀随机变量 $X_1, \cdots, X_n$ 的 n 个次序统计量, $X_{(i)}$ 的密度为

$$f(x) = \frac{n!}{(i-1)!(n-i)!} x^{i-1}(1-x)^{n-i} \qquad 0 < x < 1$$

(a) 计算 $\operatorname{Var}(X_{(i)}), i = 1, \cdots, n$. (b) 找出使得 $\operatorname{Var}(X_{(i)})$ 达最小和最大值的 i 的值.

22. 设 $Y = a + bX$, 证明

$$\rho(X,Y) = \begin{cases} +1 & b > 0 \\ -1 & b < 0 \end{cases}$$

23. 设 Z 为标准正态随机变量, $Y = a + bZ + cZ^2$, 证明

$$\rho(Y,Z) = \frac{b}{\sqrt{b^2 + 2c^2}}$$

24. 证明柯西–施瓦兹不等式, 即

$$(E[XY])^2 \leqslant E[X^2]E[Y^2]$$

提示: 当 $Y = tX$ 对某一 t 成立时, 上式等号成立. 其他情况下,

$$0 < E[(tX+Y)^2] = E[X^2]t^2 + 2E[XY]t + E[Y^2]$$

因此, 关于 t 的二次方程

$$E[X^2]t^2 + 2E[XY]t + E[Y^2] = 0$$

的根为复数, 故其判别式为负值.

25. 设 X, Y 相互独立, 在下列两种情形下, 证明

$$E[X|Y=y] = E[X] \qquad \text{对一切}y\text{成立}$$

(a) 离散情形; (b) 连续情形.

26. 证明 $E[g(X)Y|X] = g(X)E[Y|X]$.

27. 证明: 若 $E[Y|X=x] = E[Y]$ 对一切 x 成立, 则 X, Y 不相关. 同时, 给出一个反例, 说明逆命题不真.

提示: 证明并利用公式 $E[XY] = E[XE[Y|X]]$.

28. 证明 $\mathrm{Cov}(X, E[Y|X]) = \mathrm{Cov}(X,Y)$.

29. 设 $X_1, \cdots, X_n$ 为独立同分布随机变量序列, 求

$$E[X_1|X_1 + \cdots + X_n = x]$$

30. 考虑关于多项分布的例 4f, 利用条件期望计算 $E[N_iN_j]$, 再利用此公式验证关于 $\mathrm{Cov}(N_iN_j)$ 的公式.

31. 设在坛子中, 原有 b 个黑球和 w 个白球, 每一步放入 r 个黑球然后再随机的拿出 r 个球, 指出, 经过 t 步以后

$$E[\text{白球数}] = \left(\frac{b+w}{b+w+r}\right)^t w$$

32. 令 I_A 为事件 A 的示性变量, 即

$$I_A = \begin{cases} 1, & A \text{ 在试验中发生} \\ 0, & A \text{ 在试验中不发生} \end{cases}$$

对于任意随机变量 X, 证明公式

$$E[X|A] = \frac{E[XI_A]}{P(A)}$$

33. 设掷一枚硬币, 其正面朝上的概率为 p. 现在连续掷硬币, 直到连续出现 r 次正面朝上为止, 求掷硬币次数的期望值.

提示: 记第一次出现反面朝上的时刻为 T. 在 $T=t$ 的条件下, 求掷硬币次数的期望, 得到方程[1]

$$E[X]=\sum_{t=1}^{r}E[X|T=t]\cdot P\{T=t\}+rp^r$$
$$=(1-p)\sum_{t=1}^{r}p^{t-1}(t+E[X])+rp^r$$

然后解出 $E[X]$.

34. 理论习题 33 的另一种解法. 记 T_r 表示出现连续 r 个正面朝上所需的掷硬币次数.

(a) 求 $E[T_r|T_{r-1}]$ (b) 将 $E(T_r)$ 表示成 $E[T_{r-1}]$ 的函数

(c) $E[T_1]$ 是多少? (d) $E[T_r]$ 是多少?

35. 设 X 为取非负整数值的离散型随机变量, 其分布列为 $P\{X=j\}=p_j, j\geqslant 0$, 它的概率矩母函数由下式定义:

$$\phi(s)=E[s^X]=\sum_{j=0}^{\infty}p_js^j$$

设 Y 具有几何分布, 参数 $p=1-s$, 其中 $s\in(0,1)$. 又设 Y 与 X 相互独立. 指出

$$\phi(s)=P\{X<Y\}$$

36. 设某坛子内有 a 个白球, b 个黑球. 一次从坛子内随机取出一个球, 直至坛子中的球变成同一颜色. 记 M_{ab} 表示试验结束时坛子中的球的个数的期望值. 导出一个递推公式, 并且当 $a=3, b=5$ 时求出 M_{ab}.

37. 一个坛子里有 a 个白球, b 个黑球. 从坛子里随机地取出一个球, 如果这个球是白球, 就放回坛子, 如果是黑球, 就放入一个白球作为替换. 记 M_n 表示经过 n 次取球以后坛子里的白球数的期望值.

(a) 导出下面的递推公式

$$M_{n+1}=\Big(1-\frac{1}{a+b}\Big)M_n+1$$

(b) 利用 (a) 证明

$$M_n=a+b-b\Big(1-\frac{1}{a+b}\Big)^n$$

(c) 第 $n+1$ 次从坛子里取出一个白球的概率是多少?

38. 设 Y, X_1, X_2 为随机变量. 基于 X_1, X_2 的 Y 的最优线性预测是使

$$E[(Y-(a+bX_1+cX_2))^2]$$

达到最小的 $a+bX_1+cX_2$. 求最优线性预测中的系数 a,b,c.

39. Y 的基于 X 的最优二次预测 $a+bX+cX^2$ 是使

$$E[(Y-(a+bX+cX^2))^2]$$

达到最小的二次三项式 $a+bX+cX^2$. 求最优二次预测中的系数 a,b,c 的值.

---

[1] 此方程与原文稍有不同. —— 译者注

40. 利用条件方差公式确定参数为 p 的几何随机变量 X 的方差.

41. 设 X 为正态随机变量, 其参数 $\mu = 0, \sigma^2 = 1$, 令 I 为与 X 相互独立的随机变量, $P\{I=1\} = \frac{1}{2} = P\{I=0\}$. Y 由下式定义:

$$Y = \begin{cases} X & I=1 \\ -X & I=0 \end{cases}$$

即 Y 以相同的可能等于 X 或 $-X$.

(a) X, Y 是否相互独立? (b) I, Y 是否相互独立?

(c) 指出 Y 的分布为标准正态分布. (d) 指出 $\mathrm{Cov}(X,Y) = 0$.

42. 设 Y 的依赖于 X 的最优线性预测为 $\mu_y + \rho\frac{\sigma_x}{\sigma_y}(X - \mu_x)$, 则由命题 6.1 可知, 若

$$E[Y|X] = a + bX$$

则

$$a = \mu_y - \rho\frac{\sigma_y}{\sigma_x}\mu_x \qquad b = \rho\frac{\sigma_y}{\sigma_x}$$

(为什么?) 请直接证明此结论.

43. 对于随机变量 X 和 Z, 证明

$$E[(X-Y)^2] = E[X^2] - E[Y^2]$$

其中 $Y = E[X|Z]$.

44. 考虑一个总体, 总体中的每一个个体都能产生后代. 假定总体中的每一个个体, 当它的生命结束的时候, 具有 j 个子代的概率为 $P_j, j \geqslant 0$, 子代的个数与别的个体的子代个数独立. 现设最早的个体的个数 X_0 称为第 0 代的大小. 第 0 代个体的子代称为第 1 代个体, 记第 1 代个体总数为 X_1, 依此类推, 可得到第 n 代个体总和为 X_n, 记 $\mu = \sum_{j=0}^{\infty} jP_j$, $\sigma^2 = \sum_{j=0}^{\infty}(j-\mu)^2 P_j$, 它们分别表示一个个体的子代的个数的期望和方差. 现假定 $X_0 = 1$, 即第 0 代个体只有一个.

(a) 指出 $E[X_n] = \mu E[X_{n-1}]$. (b) 利用 (a), 证明 $E[X_n] = \mu^n$.

(c) 指出 $\mathrm{Var}(X_n) = \sigma^2\mu^{n-1} + \mu^2\mathrm{Var}(X_{n-1})$.

(d) 利用 (c), 证明

$$\mathrm{Var}(X_n) = \begin{cases} \sigma^2\mu^{n-1}\left(\frac{\mu^n-1}{\mu-1}\right) & \mu \neq 1 \\ n\sigma^2 & \mu = 1 \end{cases}$$

刚才介绍的模型称为分枝过程. 分枝过程的一个重要问题是总体灭绝的概率, 记 π 表示由一个个体出发, 其总体灭绝的概率, 即

$$\pi = P\{\text{总体灭绝}|X_0 = 1\}.$$

(e) 证明: π 满足

$$\pi = \sum_{j=0}^{\infty} P_j \pi^j$$

提示: 以第 1 代的个体数作条件, 求出灭绝的条件概率.

45. 验证表 7.2 中均匀分布的矩母函数的公式, 同时用求微商方法求出相应随机变量的期望和方差.

46. 对于标准正态随机变量 Z, 令 $\mu_n = E[Z^n]$, 指出

$$\mu_n = \begin{cases} 0 & n\text{为奇数} \\ \dfrac{(2j)\,!}{2^j j\,!} & n = 2j \end{cases}$$

提示: 将 Z 的矩母函数在 0 处展成泰勒级数:

$$E[e^{tZ}] = e^{t^2/2} = \sum_{j=0}^{\infty} \frac{(t^2/2)^j}{j\,!}$$

47. 设 X 是一个正态随机变量, 其参数为 (μ, σ^2). 利用理论习题 46 证明

$$E[X^n] = \sum_{j=0}^{[n/2]} \frac{\binom{n}{2j}\mu^{n-2j}\sigma^{2j}(2j)\,!}{2^j j\,!}$$

其中, $[n/2]$ 表示 $n/2$ 的整数部分. 在 $n = 1, 2$ 的情况下, 通过直接计算验证本题的公式.

48. 设 $Y = aX + b$, 其中 a, b 为常数. 将 Y 的矩母函数用 X 的矩母函数表达出来.

49. 取正值的随机变量 X 称为对数正态随机变量, 参数为 (μ, σ^2). 如果 $\ln X$ 的分布为正态, 期望为 μ, 方差为 σ^2. 利用正态矩母函数找出对数正态随机变量的期望和方差.

50. 设 X 的矩母函数为 $M(t)$, 定义 $\Psi(t) = \ln M(t)$. 证明 $\Psi''(t)|_{t=0} = \mathrm{Var}(X)$.

51. 利用表 7.2, 求出 $\sum_{i=1}^{n} X_i$ 的分布, 其中 X_i 为独立同分布的随机变量, 其公共分布为指数分布, 期望为 $1/\lambda$.

52. 从 X, Y 的联合矩母函数求出 $\mathrm{Cov}(X, Y)$.

53. 设 $X_1, \cdots, X_n$ 具有多元正态分布. 指出 $X_1, \cdots, X_n$ 相互独立的充要条件是

$$\mathrm{Cov}(X_i, X_j) = 0, i \neq j$$

54. 设 Z 为标准正态随机变量, 计算 $\mathrm{Cov}(Z, Z^2)$.

55. 设 Y 具有正态分布, 其均值为 μ, 方差为 σ^2. 假设 $Y = y$ 的条件下 X 的分布为正态分布, 期望为 y, 方差为 1.
   (a) 指出 (X, Y) 的联合分布与 $(Y + Z, Y)$ 的联合分布相同, 其中 Z 与 Y 相互独立, 且具有标准正态分布.
   (b) 利用 (a), 证明 (X, Y) 具有二元正态分布.　(c) 求 $E[X], \mathrm{Var}(X), \mathrm{Corr}(X, Y)$.
   (d) 求 $E[Y|X = x]$.　(e) Y 在 $X = x$ 之下的条件分布是什么?

自 检 习 题

1. 设由 m 个名称组成的列表, 在列表上同一名称可出现好几次. 记 $n(i)$ 表示在第 i 个位置上的名称在表上出现的次数 $i = 1, \cdots, m$, 用 d 表示列表上不同名称的个数.

(a) 将 d 表成 $m, n(i), i=1,\cdots,m$ 的函数. 令 U 为 $(0,1)$ 上均匀分布的随机变量, 令 $X=[mU]+1$.

(b) X 的分布列是什么?

(c) 指出 $E[m/n(X)]=d$.

2. 设坛子里有 n 个白球, m 个黑球, 随机地从坛子中将球一个一个地取出. 数一数取出一个黑球后紧接着取出一个白球的次数, 并求其期望值.

3. 10 对夫妇被安排在 5 张餐桌上, 每张餐桌上有 4 个座位.

(a) 如果座位是完全随机地安排的, 求坐在同一张餐桌的夫妇的对数的期望值.

(b) 如果随机地挑选 2 个男人, 2 个女人坐在一张桌子上, 求坐在同一桌的夫妇的对数的期望值.

4. 设连续掷一枚骰子, 一直到 6 个面都出现为止. 求点数“1”出现次数的期望值.

5. 设有一副牌, 由 n 张红牌和 n 张黑牌组成. 将牌洗好以后, 顺次地一张一张翻开, 当翻出一张红牌并且此时已翻开的红牌数比黑牌数多时, 你赢得 1 个单位.(例如, 若 $n=2$, 翻牌结果是 rbrb, 则此时你总共赢得 2 个单位) 求你赢得的单位数的期望值.

6. 设 $A_1, A_2, \cdots, A_n$ 为 n 个事件, 记 N 为这些事件的发生数, I 为一随机变量,

$$I=\begin{cases}1 & \text{如果所有事件都发生}\\ 0 & \text{其他}\end{cases}$$

证明下面的 Bonferroni 不等式:

$$P(A_1\cdots A_n)\geqslant\sum_{i=1}^{n}P(A_i)-(n-1)$$

提示: 首先证明 $N\leqslant n-1+I$.

7. 设从 $\{1,2,\cdots,n\}$ 中随机地取 k 个数. 记 X 为其中最小的数, 将 X 解释为服从负超几何分布的随机变量, 然后求 $E[X]$ 的值.

8. 一架飞机载着 r 个家庭着陆, 有 n_j 家带有 j 件行李, $\sum\limits_j n_j=r$. 当飞机着陆以后, $N=\sum\limits_j jn_j$ 件行李按随机的顺序一件一件的取出来. 一旦一个家庭收齐了他们的行李就立刻离开机场. 现设桑切斯一家有 j 件行李, 求在桑切斯一家之后离开机场的家庭数的期望值.

***9.** 在半径为 1 的圆周上, 放上 19 件器材. 指出无论怎么安放这 19 个器材, 都至少存在一段长度为 1 的弧, 其上至少包含 4 个器件.

10. 设 X 是泊松随机变量, 均值为 λ. 指出, 若 λ 不是太小, 有下式成立:

$$\mathrm{Var}(\sqrt{X})\approx 0.25$$

提示: 利用理论习题 4 中的结果去近似 $E[\sqrt{X}]$.

11. 设在自检习题 3 中 20 个人安排了 7 张桌子, 其中 3 张桌子上有 4 个位子, 4 张桌子上有 2 个位子. 如果 20 个人是随机地坐, 求坐在同一桌的夫妇对数的期望值.

12. 设员工 $1, \cdots, n$, 被招聘到某公司, 他们是这样被招聘进来的, 由第 1 个人开公司, 将 2 招聘进来之后, 1 和 2 就竞争着招聘 3. 当 3 被招聘进来以后, 1, 2, 3 就竞争着招聘 4. 假定当 $1, \cdots, i$ 竞争着招聘 $i+1$ 时, 这 i 个人中每一个人能招聘到 $i+1$ 的概率是相等的.

(a) 求 $1, 2, \cdots, n$ 中没有招聘到其他人的人数的期望值.

(b) 求出没有招聘到其他人的人数的方差. 对于 $n = 5$, 求出它的值.

13. 9 个人组成一个篮球队, 其中 2 个中锋, 3 个前锋, 4 个后卫. 现将 9 个人随机地分成 3 组. 一组称为完全的, 如果这个组含有一个中锋, 一个前锋, 一个后卫. 求完全组个数的

(a) 期望值. (b) 方差.

14. 从一副 52 张的扑克牌中随机地抽出 13 张, 分别记 X 和 Y 为“A”的张数和黑桃的张数.

(a) 指出 X, Y 不相关. (b) 它们独立吗?

15. 设在箱子内有一批硬币, 每一个硬币有一个 p 值, 当抛掷这枚硬币时, 正面朝上的概率为 p, 当从箱子内取出一硬币时, 硬币的 p 值是在 $(0,1)$ 上均匀分布的. 现在, 设硬币已经取出, 在抛掷以前, 你必须猜一下抛掷结果, 猜对了会赢一个单位, 反之会输一个单位.

(a) 若不告诉你 p 的值, 你的期望所得是多少?

(b) 若在猜测以前, 你可以调查得到 p 的值, 你应该作怎样的猜测?

(c) 计算 (b) 中你的期望所得.

16. 在自检习题 1 中, 我们指出可以利用 $(0,1)$ 上的均匀随机变量 (通常称随机数), 得到一个随机变量, 且其均值刚好等于列表上不同名称的个数. 但是, 那里的方法要求选择一个随机的位置, 并且确定该位置上的名称在列表上出现的次数, 现在提供另一个方法, 这个方法在名称重复次数较多时较为有效. 其方法如下: 首先像习题 1 中那样选定一个随机变量 X, 然后确认在位置 X 上的名称, 再从列表的开头开始检查, 直到这个名称出现为止. 如果这个名称出现在 X 的前面, 则定义 $I = 0$, 否则定义 $I = 1$, 证明 $E[mI] = d$.

提示: 利用条件期望计算 $E[I]$.

17. 一共有 m 个物件, 顺次放入 n 个房间, 每一个物件放在房间 j 的概率为 $p_j, j = 1, 2, \cdots, n$. 当某一物件放入一非空的房间, (该房间已经被别的物体占据着) 时, 则称为出现一个碰撞. 求碰撞数的期望值.

18. 设 n 个 1 和 m 个 0 随机地排成一列, 令 X 表示这个排列的第一个游程的长度.(既可以是“1”的游程, 也可以是“0”的游程. 例如, 00101, 此时 $X = 2$.) 求 $E[X]$.

19. 盒子 H 内有 n 个物件, 盒子 T 内有 m 个物件. 有一硬币, 抛掷时以概率 p 正面朝上, 以概率 $1-p$ 反面朝上. 当正面朝上时, 从盒子 H 中取走一个物件, 当反面朝上时, 从盒子 T 中取走一个物体. 当有一个盒子为空的时候, 例如, H 为空的时候, 而硬币此时又是正面朝上, 只好不拿物件. 而继续下一次掷硬币, 这个过程一直继续到两个盒子全空为止. 求在两个盒子全空时, 所掷硬币次数的期望值.

提示: 以前 $n+m$ 次抛掷的正面朝上数为条件.

20. 设 X 为非负随机变量, 其分布函数为 $F(x)$. 记 $\bar{F}(x) = 1 - F(x)$. 证明:

$$E[X^n] = \int_0^\infty x^{n-1}\bar{F}(x)\mathrm{d}x$$

提示：利用

$$X^n = n\int_0^X x^{n-1}\mathrm{d}x = n\int_0^\infty x^{n-1}I_X(x)\mathrm{d}x$$

其中

$$I_X(x) = \begin{cases} 1 & x < X \\ 0 & \text{其他} \end{cases}$$

***21.** 设 $a_1, \cdots, a_n$ 不全为 0, $\sum_{i=1}^n a_i = 0$, 指出存在一个排列 $i_1, \cdots, i_n$, 使得

$$\sum_{j=1}^n a_{i_j}a_{i_{j+1}} < 0$$

提示：利用概率化方法.(很有意思的是, 可以不存在排列, 使得连续两项的乘积之和为正, 例如 $n = 3, a_1 = a_2 = -1, a_3 = 2$, 此时, 对所有排列, 其相应的连续两项的乘积和均小于或等于 0.)

22. 设 $X_i(i = 1,2,3)$ 为相互独立的泊松随机变量, 均值分别为 $\lambda_i(i = 1,2,3)$. 令 $X = X_1 + X_2, Y = X_2 + X_3$, (X, Y) 称为具有二维泊松分布的随机向量. 求：

(a) $E[X]$ 和 $E[Y]$; (b) $\mathrm{Cov}(X, Y)$; (c) (X, Y) 的联合分布列 $P\{X = i, Y = j\}$.

23. 令 $(X_i, Y_i), i = 1, 2, \cdots$ 为独立同分布的随机向量序列, 即 (X_1, Y_1) 与 (X_2, Y_2) 相互独立且同分布, 等等. 虽然, X_i, Y_i 不独立, 但 $X_i, Y_j, i \neq j$ 却相互独立, 令

$$\mu_x = E[X_i], \mu_y = E[Y_i], \sigma_x^2 = \mathrm{Var}(X_i), \sigma_y^2 = \mathrm{Var}(Y_i), \rho = \mathrm{Corr}(X_i, Y_i)$$

求 $\mathrm{Corr}(\sum_{i=1}^n X_i, \sum_{j=1}^n Y_i)$.

24. 从一副 52 张牌内抽取 3 张牌 (无放回), 记 X 表示选中的“A”的张数.

(a) 求 $E[X|$黑桃“A”已选中$]$. (b) 求 $E[X|$至少一张“A”已选中$]$.

25. 记 Φ 为标准正态分布函数, X 为正态随机变量, 其均值为 μ, 方差 $\sigma^2 = 1$. 我们要计算 $E[\Phi(X)]$. 为此, 令 Z 为标准正态随机变量, 且与 X 相互独立, 令

$$I = \begin{cases} 1 & Z < X \\ 0 & Z \geqslant X \end{cases}$$

(a) 指出 $E[I|X = x] = \Phi(x)$. (b) 指出 $E[\Phi(X)] = P\{Z < X\}$. (c) 指出 $E[\Phi(X)] = \Phi\left(\mu/\sqrt{2}\right)$.

提示 (c)：$X - Z$ 的分布是什么?

这个题目源自统计学. 我们将观察随机变量 X, X 的分布为正态分布, 期望为 μ, 方差为 1, 我们希望检验假设 $\mu \geqslant 0$. 显然, 当 X 充分小的时候, 拒绝 $\mu \geqslant 0$ 这个假设. 若 $X = x$, 则这个假设的 p 值定义为 $\mu = 0$ 的假定之下随机事件 $\{X \leqslant x\}$ 的概率 (当 p 值小的时候, 说明原来的假设可能是假的). 由于当 $\mu = 0$ 时, X 具有标准正态分布. 因此, p 值为 $\Phi(x)$. 当 μ 为真值时, p 值的平均值为 $\Phi\left(\frac{\mu}{\sqrt{2}}\right)$.

26. 设有一硬币, 抛掷时以 p 的概率正面朝上. 现在设连续抛掷这枚硬币, 直到出现了 n 次正面朝上或者 m 次反面朝上为止, 求抛掷硬币次数的期望值.

提示: 设想, 当达到目的以后, 还继续掷硬币. 令 X 表示为得到 n 次正面朝上所需的掷硬币次数, Y 表示为得到 m 次反面朝上所需的掷硬币次数. 注意 $\max(X,Y)+\min(X,Y)=X+Y$. 为计算 $E[\max(X,Y)]$, 首先计算在头 $n+m-1$ 次掷硬币中正面朝上次数固定的条件下 $\max(X,Y)$ 的条件期望.

27. 设有一副牌共有 n 张, 一个洗牌的程序是这样的: 每次随机地抽出一张牌放在最上面. 抽出一张牌称为新牌, 指的是抽出的一张牌是以前从没有抽到过的牌. 这个洗牌过程一直到抽出 $n-1$ 张新牌就停止. 这样的过程可以使 $n!$ 种排列次序以相同的概率出现. 求出这个洗牌过程的抽牌次数的期望值.

28. 设有一独立重复试验, 每次成功的概率为 p. 当第一次试验成功或者一直到 n 次都未成功则试验停止. 求出试验次数的平均值.

提示: 利用公式 $E[X]=\sum_{i=1}^{\infty}P\{X\geqslant i\}$, 其中 X 为非负整数值随机变量.

29. 设 X,Y 均为贝努利随机变量. 指出 X,Y 相互独立的充要条件为

$$\text{Cov}(X,Y)=0.$$

30. 广义配对问题是如下叙述的. 一共有 n 个妇女, 其中有 n_i 个妇女戴的帽子的大小编号为 i, $\sum_{i=1}^{r}n_i=n$. 现在有 n 顶帽子, 大小编号为 i 的帽子有 h_i 顶, $\sum_{i=1}^{r}h_i=n$. 现在每人随机地取一顶帽子 (无放回). 找出拿到自己大小编号的帽子的人数的期望数.

第 8 章　极限定理

8.1　引　　言

在概率论中，最重要的理论结果是极限定理．极限定理中最重要的是大数定律和中心极限定理．通常，当随机变量序列的平均值在某种条件下收敛到某期望值的时候，就是大数定律．另一方面，当大量随机变量之和的分布在某种条件下逼近于正态分布时，就称为中心极限定理．

8.2　切比雪夫不等式及弱大数律

我们开始介绍马尔可夫不等式．

命题 2.1（马尔可夫不等式）　设 X 为取非负值的随机变量，则对于任何常数 $a>0$，有

$$P\{X \geqslant a\} \leqslant \frac{E[X]}{a}$$

证明：对于 $a>0$ 令

$$I=\begin{cases}1 & X \geqslant a \\ 0 & \text{其他}\end{cases}$$

由于 $X \geqslant 0$，我们有

$$I \leqslant \frac{X}{a}$$

两边求期望，得

$$E[I] \leqslant \frac{1}{a} E[X]$$

上式说明 $E[X]/a \geqslant E[I]=P\{X \geqslant a\}$，即命题结论成立．　□

作为推论，可得下列命题．

命题 2.2（切比雪夫不等式）　设 X 是一随机变量，期望 μ，方差 σ^2，则对任何 $k>0$，有

$$P\{|X-\mu| \geqslant k\} \leqslant \frac{\sigma^2}{k^2}$$

证明：由于 $(X-\mu)^2$ 为非负随机变量，利用马尔可夫不等式，得

$$P\{(X-\mu)^2 \geqslant k^2\} \leqslant \frac{E[(X-\mu)^2]}{k^2} \tag{2.1}$$

由于 $(X-\mu)^2 \geqslant k^2$ 与 $|X-\mu| \geqslant k$ 是等价的, 因此

$$P\{|X-\mu| \geqslant k\} \leqslant \frac{E[(X-\mu)^2]}{k^2} = \frac{\sigma^2}{k^2} \qquad \square$$

马尔可夫不等式和切比雪夫不等式的重要性在于: 当我们只知道随机变量的期望, 或期望和方差都知道时, 可以导出概率的上界. 当然, 我们知道概率分布时, 就可以直接计算概率的值而不必计算概率的上界.

例 2a 某工厂在一周内生产的产品的件数为随机变量, 假定已知这个随机变量的期望值为 50.

(a) 本周内产品超过 57 件的概率有多大?

(b) 如果我们进一步知道每周产量的方差为 25, 那么本周产量在 40 到 60 之间的概率有多大?

解: 记 X 为本周的产量.

(a) 由过去的经验只知道 X 的期望为 $\mu = 50$, 而不知道 X 的分布. 因此, 我们不能计算相应的概率值, 只能利用不等式计算出相应概率的界. 利用马尔可夫不等式, 得

$$P\{X > 75\} \leqslant \frac{E[X]}{75} = \frac{50}{75} = \frac{2}{3}$$

(b) 基于与 (a) 一样的原因, 我们利用切比雪夫不等式

$$P\{|X-50| \geqslant 10\} \leqslant \frac{\sigma^2}{10^2} = \frac{1}{4}$$

故

$$P\{|X-50| < 10\} \geqslant 1 - \frac{1}{4} = \frac{3}{4}$$

故本周内的产品在 40 到 60 之间的概率至少为 0.75. ■

由于切比雪夫不等式适用于所有的分布, 因此, 不能指望所得的概率的界与真实的概率很接近. 下面看一个例子.

例 2b 设 X 为 $(0,10)$ 上具有均匀分布的随机变量, 已知 $E[X]=5, \mathrm{Var}(X) = 25/3$. 利用切比雪夫不等式可得

$$P\{|X-5| > 4\} \leqslant \frac{25}{3 \times 16} \approx 0.52$$

而实际上, 这个概率为

$$P\{|X-5| > 4\} = 0.20$$

由上式看出, 我们只能利用切比雪夫不等式找到概率的界, 但不能用来估计概率值本身.

类似地, 设 X 具有正态分布, 期望为 μ, 方差为 σ^2, 利用切比雪夫不等式得到

$$P\{|X-\mu|>2\sigma\} \leqslant \frac{1}{4}$$

而实际概率为

$$P\{|X-\mu|>2\sigma\} = P\left\{\left|\frac{X-\mu}{\sigma}\right|>2\right\} = 2[1-\Phi(2)] \approx 0.0456$$

两者相差很远. ■

切比雪夫不等式的主要用途是证明理论结果, 例如命题 2.3, 但是最重要的是证明大数定律.

命题 2.3 若 $\mathrm{Var}(X)=0$, 则 $P\{X=E[X]\}=1$. 也即是说, 一个随机变量的方差为 0 的充要条件是这个随机变量以概率为 1 地等于常数.

证明: 利用切比雪夫不等式, 对任何 $n \geqslant 1$

$$P\left\{|X-\mu|>\frac{1}{n}\right\} = 0$$

令 $n \to \infty$, 得

$$0 = \lim_{n\to\infty} P\left\{|X-\mu|>\frac{1}{n}\right\} = P\left\{\lim_{n\to\infty}\left\{|X-\mu|>\frac{1}{n}\right\}\right\} = P\{X \neq \mu\}$$

结论得到证明. □

定理 2.1 (弱大数律) 设 $X_1, X_2, \cdots$ 为独立同分布的随机变量序列, 其公共期望 $E[X_i]=\mu$ 为有限, 则对任何 $\varepsilon>0$,

$$P\left\{\left|\frac{X_1+\cdots+X_n}{n}-\mu\right| \geqslant \varepsilon\right\} \to 0 \qquad n\to\infty$$

证明: 我们只在 $\mathrm{Var}(X_i)=\sigma^2$ 为有限的情形下证明此定理. 此时

$$E\left[\frac{X_1+\cdots+X_n}{n}\right] = \mu, \mathrm{Var}\left(\frac{X_1+\cdots+X_n}{n}\right) = \frac{\sigma^2}{n}$$

利用切比雪夫不等式, 得

$$P\left\{\left|\frac{X_1+\cdots+X_n}{n}-\mu\right| \geqslant \varepsilon\right\} \leqslant \frac{\sigma^2}{n\varepsilon^2}$$

由上式看出, 定理显然成立. □

弱大数律最早是由詹姆士 · 伯努利证明的. 他证明的大数律是一种特殊情况, 其中 X_i 只取 0 或 1, 即 X 为伯努利随机变量. 他对该定理的陈述和证明见于他的一本书《推测术》, 这本书出版于 1713 年, 是在詹姆士 · 伯努利去世后 8 年, 由他的同为数学家的侄子尼古拉斯 · 伯努利整理出版. 要知道, 当时切比雪夫不等式还

不为人知, 伯努利必须借助十分灵巧的方法证明其结果. 定理 2.1 是独立同分布序列的大数律的最一般形式, 它由前苏联数学家辛钦所证明.

8.3 中心极限定理

中心极限定理是概率论中最著名的结果之一, 用粗略的语言来说, 大量的独立随机变量之和的分布近似地为正态分布. 因此, 中心极限定理为计算独立随机变量和的有关概率提供了理论依据, 同时也解释了现实世界中许多实际的总体的分布的频率曲线呈现钟形曲线 (即正态的) 的原因.

下面叙述的是最简单的中心极限定理.

定理 3.1 (中心极限定理) 设 $X_1, X_2, \cdots$ 为独立同分布序列, 其公共分布的期望为 μ, 方差为 σ^2. 则随机变量

$$\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

的分布当 $n \to \infty$ 时趋向于标准正态分布. 即对任何 $a \in (-\infty, \infty)$,

$$P\left\{\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leqslant a\right\} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} \mathrm{e}^{-x^2/2} \mathrm{d}x \qquad n \to \infty$$

证明的关键之处是下面的一条引理, 由于证明涉及太多数学上的细节, 我们只给出陈述.

引理 3.1 设 $Z_1, Z_2, \cdots$ 为一随机变量序列, 其分布函数为 F_{Z_n}, 相应的矩母函数为 $M_{Z_n}, n \geqslant 1$. 又设 Z 的分布为 F_Z, 矩母函数为 M_Z, 若 $M_{Z_n}(t) \to M_Z(t)$ 对一切 t 成立, 则 $F_{Z_n}(t) \to F_Z(t)$ 对 $F_Z(t)$ 所有的连续点成立.

若 Z 为标准正态随机变量, 则 $M_z(t) = \mathrm{e}^{t^2/2}$, 利用引理 3.1 可知, 若 $M_{Z_n}(t) \to \mathrm{e}^{t^2/2}, n \to \infty$, 则 $F_{Z_n}(t) \to \Phi(t), n \to \infty$.

现在证明中心极限定理.

中心极限定理的证明: 首先, 假定 $\mu = 0, \sigma^2 = 1$, 我们只在 X_i 的矩母函数 $M(t)$ 存在且有限的假定之下证明定理. 现在, $X_i/\sqrt{n}$ 的矩母函数为

$$E\left[\exp\left\{\frac{tX_i}{\sqrt{n}}\right\}\right] = M\left(\frac{t}{\sqrt{n}}\right)$$

由此可知, $\sum_{i=1}^{n} X_i/\sqrt{n}$ 的矩母函数为 $\left[M\left(\frac{t}{\sqrt{n}}\right)\right]^n$. 记

$$L(t) = \ln M(t)$$

对于 $L(t)$, 我们有

$$L(0) = 0 \qquad L'(0) = \frac{M'(0)}{M(0)} = \mu = 0$$

$$L''(0) = \frac{M(0)M''(0) - [M'(0)]^2}{[M(0)]^2} = E[X]^2 = 1$$

要证明定理, 由引理 3.1, 我们必须证明 $[M(t/\sqrt{n})]^n \to e^{t^2/2}, n \to \infty$. 或等价地, $nL(t/\sqrt{n}) \to t^2/2, n \to \infty$.

下面的一系列等式说明这个极限式成立.

$$\lim_{n\to\infty} nL(t/\sqrt{n}) = \lim_{n\to\infty} \frac{-L'(t/\sqrt{n})n^{-3/2}t}{-2n^{-2}} \qquad \text{利用洛必达法则}$$

$$= \lim_{n\to\infty} \left[\frac{L'(t/\sqrt{n})t}{2n^{-1/2}}\right] = \lim_{n\to\infty} \left[-\frac{L''(t/\sqrt{n})n^{-3/2}t^2}{-2n^{-3/2}}\right] \quad \text{利用洛必达法则}$$

$$= \lim_{n\to\infty} \left[L''\left(\frac{t}{\sqrt{n}}\right)\frac{t^2}{2}\right] = \frac{t^2}{2}$$

这样, 在 $\mu = 0, \sigma = 1$ 的情况下, 定理得以证明. 对于一般情况, 只需考虑标准化随机变量序列, $X_i^* = (X_i - \mu)/\sigma$, 由于 $E[X_i^*] = 0, \mathrm{Var}(X_i^*) = 1$, 将已证得的结果应用于序列 X_i^*, 便可得一般情况的结论. □

注释 虽然定理 3.1 只说对每一个常数 a, 有

$$P\left\{\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leqslant a\right\} \to \Phi(a)$$

事实上, 这个收敛是对 a 一致的. [我们说函数 $f_n(a) \to f(a), n \to \infty$ 对 a 一致, 是说对任何 $\varepsilon > 0$, 存在 N, 使得当 $n \geqslant N$ 时, 不等式 $|f_n(a) - f(a)| < \varepsilon$ 对所有的 a 都成立.] ■

第一个中心极限定理是由棣莫弗在 1733 年左右给出证明的. 他的证明限于 X_i 为伯努利随机变量, 并且 $p = 1/2$. 后来, 拉普拉斯将这个结果推广到一般的 p 的情况. 由于二项随机变量可以看作独立同分布的伯努利随机变量之和, 因此, 此处得到的中心极限定理为第 5 章 4.1 节关于二项分布的正态逼近提供了理论依据. 拉普拉斯也发现了中心极限定理的更一般的形式, 但他的证明不严格, 事实上, 沿用他的方法也不可能严格化. 真正严格的证明是由俄国数学家李雅普洛夫在 1901~1902 年间给出的.

网站上有一个中心极限定理的模块演示及计算结果, 该模块将 n 个独立同分布随机变量之和的密度 (分布列) 演示出来. 每个随机变量只取 0, 1, 2, 3, 4 共 5 个值, 当输入分布列以及参数 n 以后, 就可以显示它们的和的分布列的形状. 图 8.1 显示了 (a) $n = 5$, (b) $n = 10$, (c) $n = 25$ 和 (d) $n = 100$ 的计算结果.

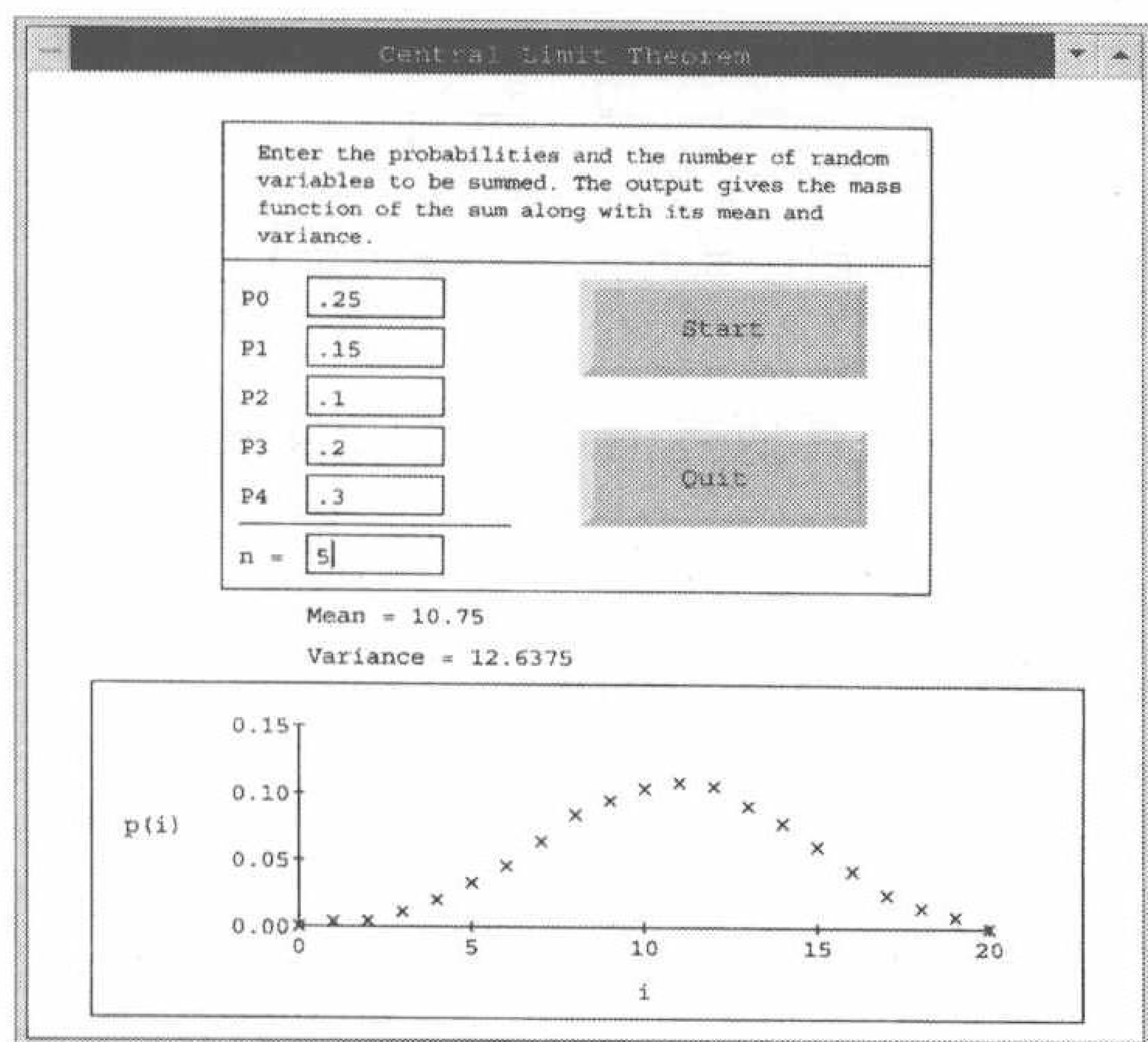


图 8.1(a) $n=5$

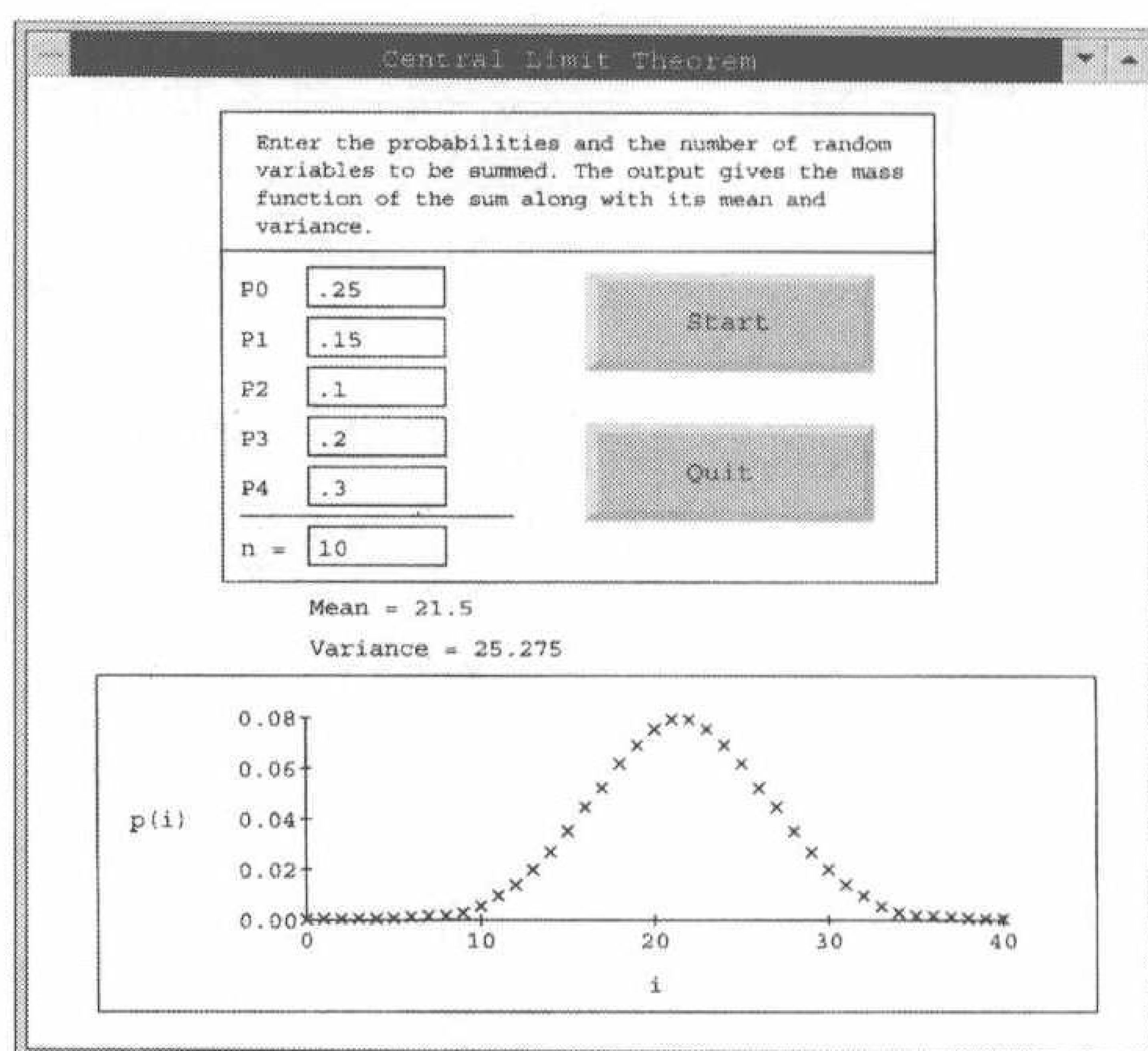


图 8.1(b) $n=10$

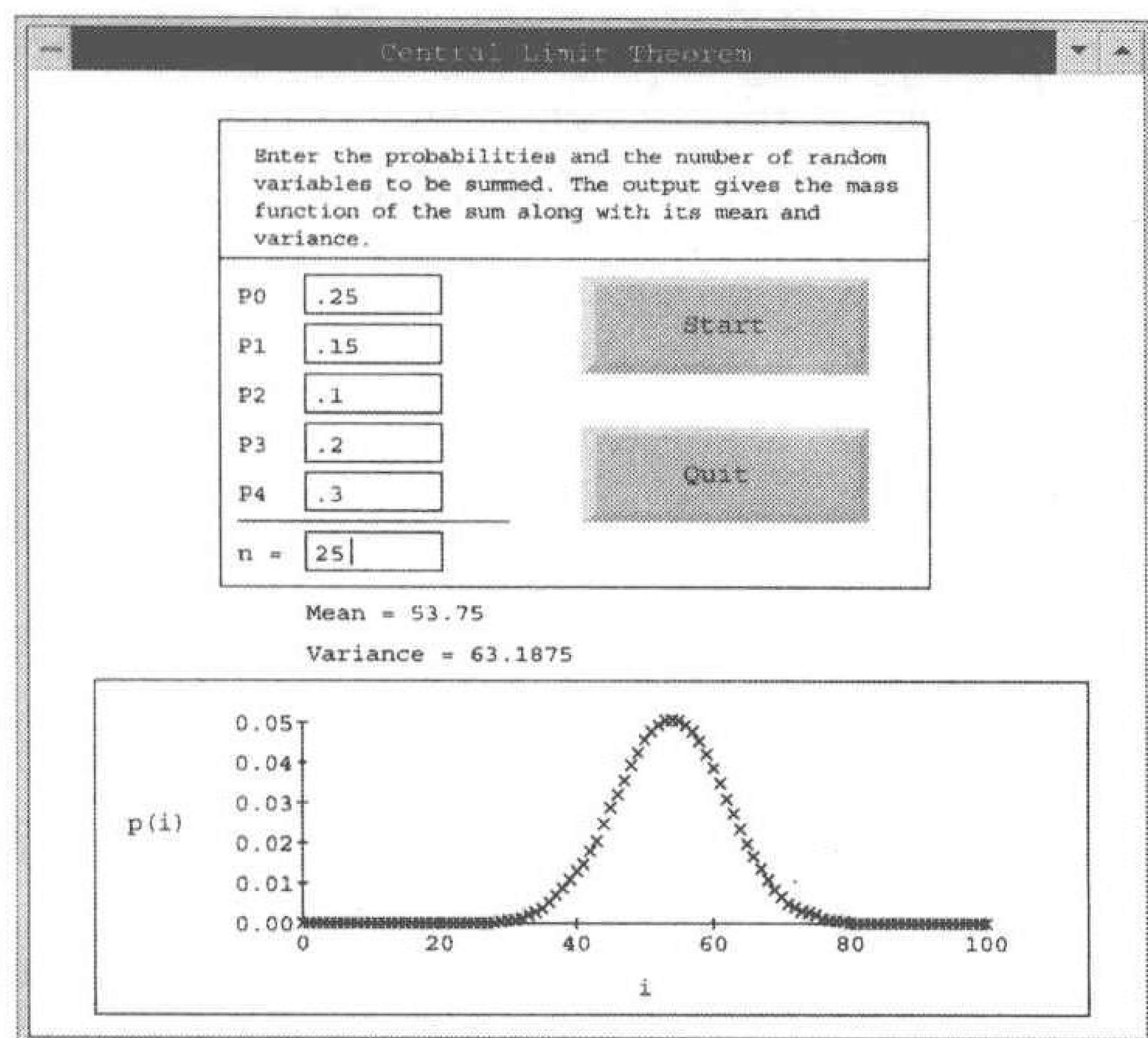


图 8.1(c) $n=25$

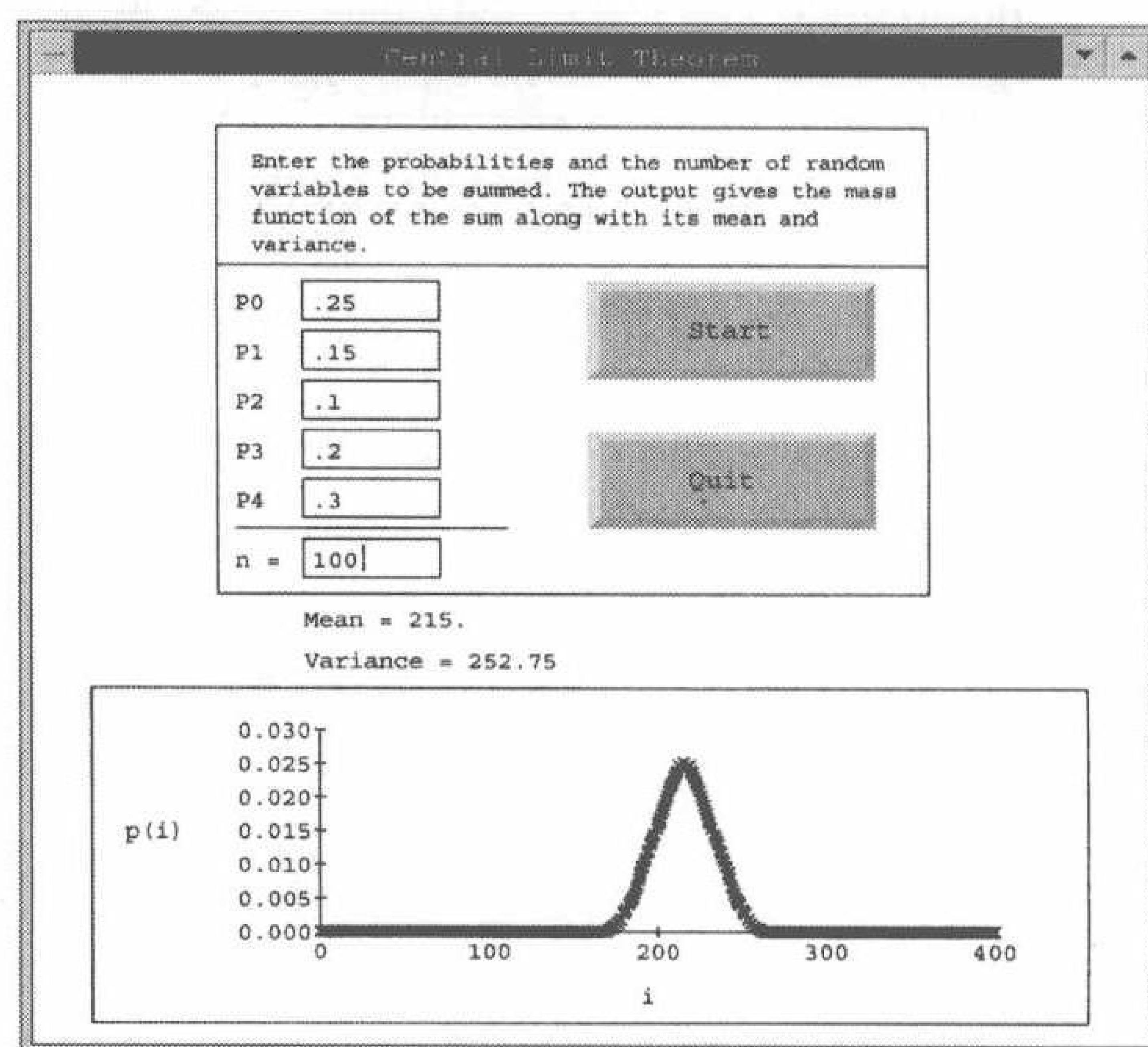


图 8.1(d) $n=100$

例 3a 一个天文学家要测量遥远的恒星到地球之间的距离 (单位: 光年). 天文学家知道, 由于大气条件以及仪器误差, 每次测量都不会得到距离的准确值, 而只是一个估计值. 因此, 天文学家只得进行一系列测量, 用这些测量值的平均值作为距离真值的估计值. 若天文学家认为各次测量值是独立同分布的各随机变量的观察值, 随机变量的公共分布的期望值为 d(距离的真值), 公共方差为 $\sigma^2=4$, 那么, 要重复测量多少次才能达到 ± 0.5 光年的精度?

解: 设观察次数为 n, $X_1,\cdots,X_n$ 为 n 次测量值, 由中心极限定理知随机变量

$$Z_n=\frac{\sum\limits_{i=1}^{n}X_i-nd}{2\sqrt{n}}$$

具有近似标准正态分布. 因此

$$P\left\{-0.5 \leqslant \frac{1}{n}\sum_{i=1}^{n} X_i - d \leqslant 0.5\right\} = P\left\{-0.5\frac{\sqrt{n}}{2} \leqslant Z_n \leqslant 0.5\frac{\sqrt{n}}{2}\right\}$$

$$\approx \Phi\left(\frac{\sqrt{n}}{4}\right) - \Phi\left(-\frac{\sqrt{n}}{4}\right) = 2\Phi\left(\frac{\sqrt{n}}{4}\right) - 1$$

如果天文学家希望以 95% 的把握保证估计值与真值之差在 0.5 光年以内, 他应作 n^* 次以上重复测量, n^* 满足

$$2\Phi\left(\frac{\sqrt{n^*}}{4}\right) - 1 = 0.95 \quad 或 \quad \Phi\left(\frac{\sqrt{n^*}}{4}\right) = 0.975$$

由第 5 章的表 5.1 得

$$\frac{\sqrt{n^*}}{4} = 1.96 \quad 或 \quad n^* = (7.84)^2 \approx 61.47$$

由于 n^* 不是整数, 因此他应作 62 次重复观测.

前面的分析中有一假定, 正态逼近是好的近似. 尽管 $n = 62$ 在通常情形下, Z_n与标准正态分布已经很靠近, 但是 Z_n 与标准正态分布逼近程度还依赖于 X_i 的分布. 若天文学家对于正态逼近还没有把握, 他可以利用切比雪夫不等式. 由于

$$E\left[\sum_{i=1}^{n}\frac{X_i}{n}\right] = d \quad \mathrm{Var}\left(\sum_{i=1}^{n}\frac{X_i}{n}\right) = \frac{4}{n}$$

由切比雪夫不等式知

$$P\left\{\left|\sum_{i=1}^{n}\frac{X_i}{n} - d\right| > 0.5\right\} \leqslant \frac{4}{n(0.5)^2} = \frac{16}{n}$$

令 $16/n = 0.05$, 得 $n = 320$, 此时, 他可以以 95% 的把握保证其估计精度在 0.5 光年以内. ■

例 3b 已知某心理学课上注册的学生数是一个泊松随机变量, 其期望值为 100. 任课教授决定如果注册人数为 120 人或更多时, 他将采取分班授课的形式, 否则就一个班上课. 该教授采用分班授课的概率有多大?

解: 这个概率的精确解为 $\mathrm{e}^{-100}\sum_{i=120}^{\infty}\frac{(100)^i}{i!}$, 这个数并没有给出这个概率有多大的概念. 然而, 均值为 100 的泊松随机变量可以看成 100 个均值为 1 的独立同分布的泊松随机变量之和, 由此, 可以利用中心极限定理得到近似解. 记 X 为注册的学生数

$$P\{X \geqslant 120\} = P\{X \geqslant 119.5\} \qquad 连续性修正$$

$$= P\left\{\frac{X-100}{\sqrt{100}} \geqslant \frac{119.5-100}{\sqrt{100}}\right\} \approx 1 - \Phi(1.95) \approx 0.0256$$

此处, 我们在应用中心极限定理时, 利用了泊松分布的期望和方差相等这一事实. ■

例 3c 设一共掷 10 枚均匀的骰子, 找出点数之和在 30 和 40 之间的概率的近

似值.

解: 设 X_i 表示第 i 个骰子的值, $i=1,2,\cdots,10$. 由于

$$E(X_i)=\frac{7}{2} \quad \mathrm{Var}(X_i)=E[X_i^2]-(E[X_i])^2=\frac{35}{12}$$

利用中心极限定理

$$P\{29.5\leqslant X\leqslant 40.5\}=P\left\{\frac{29.5-35}{\sqrt{\frac{350}{12}}}\leqslant\frac{X-35}{\sqrt{\frac{350}{12}}}\leqslant\frac{40.5-35}{\sqrt{\frac{350}{12}}}\right\}$$
$$\approx 2\Phi(1.0184)-1\approx 0.692$$ ■

例 3d 令 $X_i, i=1,\cdots,10$ 是相互独立的随机变量, 其公共分布为 $(0,1)$ 上的均匀分布, 计算 $P\{\sum_{i=1}^{10}X_i>6\}$ 的近似值.

解: 由于 $E[X_i]=1/2, \mathrm{Var}(X_i)=1/12$, 利用中心极限定理

$$P\left\{\sum_{i=1}^{10}X_i>6\right\}=P\left\{\frac{\sum\limits_{i=1}^{10}X_i-5}{\sqrt{10/12}}>\frac{6-5}{\sqrt{10/12}}\right\}\approx 1-\Phi(\sqrt{1.2})\approx 0.1367$$

因此, $\sum_{i=1}^{10}X_i$ 大于 6 的可能性只有 14%. ■

例 3e 一个教师需要判 50 道试题. 评判每道题需要的时间形成一个独立同分布的随机变量序列, 其公共分布的期望为 20 分钟, 标准差为 4 分钟. 问他在 450 分钟内判完至少 25 道题的概率有多大 (近似值)?

解: 令 X_i 为判第 i 道题所需的时间, 则

$$X=\sum_{i=1}^{25}X_i$$

为判完 25 道题所需的时间. 按题意, 在 450 分钟内, 他判完这 25 道题的概率为 $P\{X\leqslant 450\}$. 利用中心极限定理求其近似值. 由

$$E[X]=\sum_{i=1}^{25}E[X_i]=25\times 20=500, \quad \mathrm{Var}(X)=\sum_{i=1}^{25}\mathrm{Var}(X_i)=25\times 16=400$$

我们得到

$$P\{X\leqslant 450\}=P\left\{\frac{X-500}{\sqrt{400}}\leqslant\frac{450-500}{\sqrt{400}}\right\}$$
$$\approx P\{Z\leqslant -2.5\}=1-\Phi(2.5)=0.006$$

上式中 Z 为标准正态随机变量. ■

当 X_i 独立, 但不一定为同分布时, 中心极限定理也成立, 其中一个 (但决不是

最一般的) 版本如下所述.

定理 3.2 (相互独立随机变量序列的中心极限定理) 设 $X_1, X_2, \cdots$ 为相互独立的随机变量序列, 相应的期望和方差分别为 $\mu_i = E[X_i], \sigma_i^2 = \mathrm{Var}(X_i)$. 若 (a) X_i 为一致有界的, 即存在 M, 使得 $P\{|X_i| < M\} = 1$ 对一切 i 成立; 且 (b) $\sum_{i=1}^{\infty} \sigma_i^2 = +\infty$, 则对一切 a,

$$P\left\{\frac{\sum\limits_{i=1}^{n}(X_i - \mu_i)}{\sqrt{\sum\limits_{i=1}^{n}\sigma_i^2}} \leqslant a\right\} \to \Phi(a) \qquad n \to \infty$$

历史注记

皮埃尔 · 西蒙 · 拉普拉斯侯爵

皮埃尔 · 西蒙 · 拉普拉斯侯爵就是我们熟悉的法国数学家拉普拉斯. 中心极限定理是由拉普拉斯提出并证明的. 他观察到测量误差 (通常认为测量误差是由大量很小的偶然误差叠加而成的) 具有正态分布. 拉普拉斯也是著名的天文学家 (他被称为法国的牛顿), 他是早期概率论与统计的理论奠基者之一, 同时积极推广概率论在日常生活中的应用. 他坚信概率论对人类具有深远意义. 他在一本名为《分析概率论》的书中说:“我们发现概率论其实就是将常识问题归结为计算. 它使我们能够精确地评价凭某种直观感受到的、往往又不能解释清楚的现象……值得注意的是, 概率论这门起源于机会游戏的科学早就应该成为人类知识最重要的组成部分……生活中那些最重要的问题绝大部分恰恰是概率论问题.”

中心极限定理的应用揭示了这样的事实, 测量误差近似地正态分布, 这个统计规律是对科学的重大贡献, 在 17,18 世纪, 中心极限定理常被称为*误差频率定律*.

Francis Galton 在他 1889 年出版的书《自然遗产》中曾说过:“我知道, 几乎没有一种理论能够像误差频率定律那样神奇, 那样贴切地体现宇宙次序. 如果古希腊人知道这个规律的话, 就一定会将它人格化或神化. 它在混乱中保持着平静, 情况越复杂、混乱, 它的主导作用就越完善. 它是最卓越的、不可思议的规律.”

8.4 强大数律

强大数律是概率论中最著名的结果, 它说明, 独立同分布随机变量序列, 前 n 个观察值的平均值以概率为 1 地收敛到分布的平均值.

定理 4.1 (强大数律) 设 $X_1, X_2, \cdots$ 为一独立同分布的随机变量序列, 其公共期望值 $\mu = E[X_i]$ 为有限, 则下式以概率为 1 地成立:

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \to \mu \qquad n \to \infty$$ ①

作为强大数律的一个应用, 设有一独立重复试验序列, 令 E 为某一事件, $P(E)$ 为事件 E 发生的概率, 令

$$X_i = \begin{cases} 1 & E\text{在第}i\text{次试验中发生} \\ 0 & \text{其他} \end{cases}$$

根据强大数律,

$$\frac{X_1 + \cdots + X_n}{n} \to E[X] = P(E) \tag{4.1}$$

$X_1 + \cdots + X_n$ 表示在前 n 次试验中, 事件 E 发生的次数, (4.1) 式说明事件 E 在前 n 次试验中发生的频率以概率为 1 地收敛到它的概率 $P(E)$.

虽然这一条定律可以不加任何条件地证明, 但是在本证明中, 我们假定 X_i 具有有限 4 阶矩, 即假定 $E[X_i^4] = K < \infty$.

强大数律的证明: 首先假定 $EX_i = \mu = 0$. 记 $S_n = \sum_{i=1}^n X_i$, 考虑

$$\begin{aligned} E[S_n^4] = & E[(X_1 + \cdots + X_n)(X_1 + \cdots + X_n) \\ & \times (X_1 + \cdots + X_n)(X_1 + \cdots + X_n)] \end{aligned}$$

将上式右边期望号内的多项式展开, 得到下列各项之和:

$$X_i^4 \quad X_i^3 X_j \quad X_i^2 X_j^2 \quad X_i^2 X_j X_k \quad \text{和} \quad X_i X_j X_k X_l$$

由于 $EX_i = 0$, 利用独立性得到

$$E[X_i^3 X_j] = E[X_i^3]E[X_j] = 0$$
$$E[X_i^2 X_j X_k] = E[X_i^2]E[X_j]E[X_k] = 0$$
$$E[X_i X_j X_k X_l] = 0$$

在展式中, X_i^4 的系数为 1, 故在 $E[S_n^4]$ 中可将所有 X_i^4 的期望合并成 $nE[X_i^4]$. 对固定的 (i,j), S_n^4 的展式中 $X_i^2 X_j^2$ 一共有 $\binom{4}{2} = 6$ 项. 因此, S_n^4 的展式中与 $X_i^2 X_j^2$ 有关的那部分为 $6\sum_{i,j} X_i^2 X_j^2$, 其中求和号是对 $\{1, 2, \cdots, n\}$ 的所有两元素组合而

① 强大数律可以表为下式:

$$P\{\lim_{n\to\infty}(X_1 + \cdots + X_n)/n = \mu\} = 1$$

求的. 因此, 它的期望为 $6\binom{n}{2}E[X_i^2X_j^2]$, 这样,

$$E[S_n^4] = nE[X_i^4] + 6\binom{n}{2}E[X_i^2X_j^2] = nK + 3n(n-1)E[X_i^2]E[X_j^2]$$

在第二个等式中, 我们再一次利用了独立性. 我们注意到

$$0 \leqslant \mathrm{Var}(X_i^2) = E[X_i^4] - (E[X_i^2])^2$$

由此可得

$$(E[X_i^2])^2 \leqslant E[X_i^4] = K$$

这样,

$$E[S_n^4] \leqslant nK + 3n(n-1)K$$

从而

$$E\left[\frac{S_n^4}{n^4}\right] \leqslant \frac{K}{n^3} + \frac{3K}{n^2}$$

由上式可知

$$E\left[\sum_{n=1}^{\infty}\frac{S_n^4}{n^4}\right] = \sum_{n=1}^{\infty}E\left[\frac{S_n^4}{n^4}\right] < \infty$$

即随机变量 $\sum_{n=1}^{\infty}S_n^4/n^4$ 的期望为有限, 说明以概率为 1 地有 $\sum_{n=1}^{\infty}S_n^4/n^4 < \infty$. (若 $\sum_{n=1}^{\infty}S_n^4/n^4$ 不是概率为 1 地有限, 则 $\sum_{n=1}^{\infty}S_n^4/n^4$ 的期望为无限) 再利用级数的性质可知, 以概率为 1 地有

$$\lim_{n\to\infty}\frac{S_n^4}{n^4} = 0$$

上式说明, 以概率为 1 地有

$$\frac{S_n}{n} \to 0 \qquad n \to \infty$$

当 $\mu = E[X_i] \neq 0$ 的时候, 我们可以化成期望为 0 的情况来处理. 由于 $E[X_i - \mu] = 0$, 利用刚才得到的结论, 可知以概率为 1 地有

$$\lim_{n\to\infty}\sum_{i=1}^{n}\frac{(X_i-\mu)}{n} = 0$$

即以概率为 1 地有

$$\lim_{n\to\infty}\sum_{i=1}^{n}\frac{X_i}{n} = \mu$$

这就是定理的结论. □

在网上有两个模块可显示强大数律, 它们考虑独立同分布随机变量序列, 每个随机变量只取 0, 1, 2, 3, 4 这 5 个值, 模块模拟 n 个这样的随机变量, 然后, 计算每

个可能值出现的频数, 以及样本均值 $\sum_{i=1}^{n} X_i/n$. 这两个不同模块的差别是微小的, 只是在图像表达上有一些不同. 在使用这些模块时, 我们只需输入 n 的值, 以及随机变量 X 的取值的概率 (分布列). 图 8.2 给出了一个具体分布列的模拟结果, 参数分别为 (a) n=100, (b) $n = 1000$, (c) n=10 000. 模拟结果显示样本均值趋于 $E[X]$.

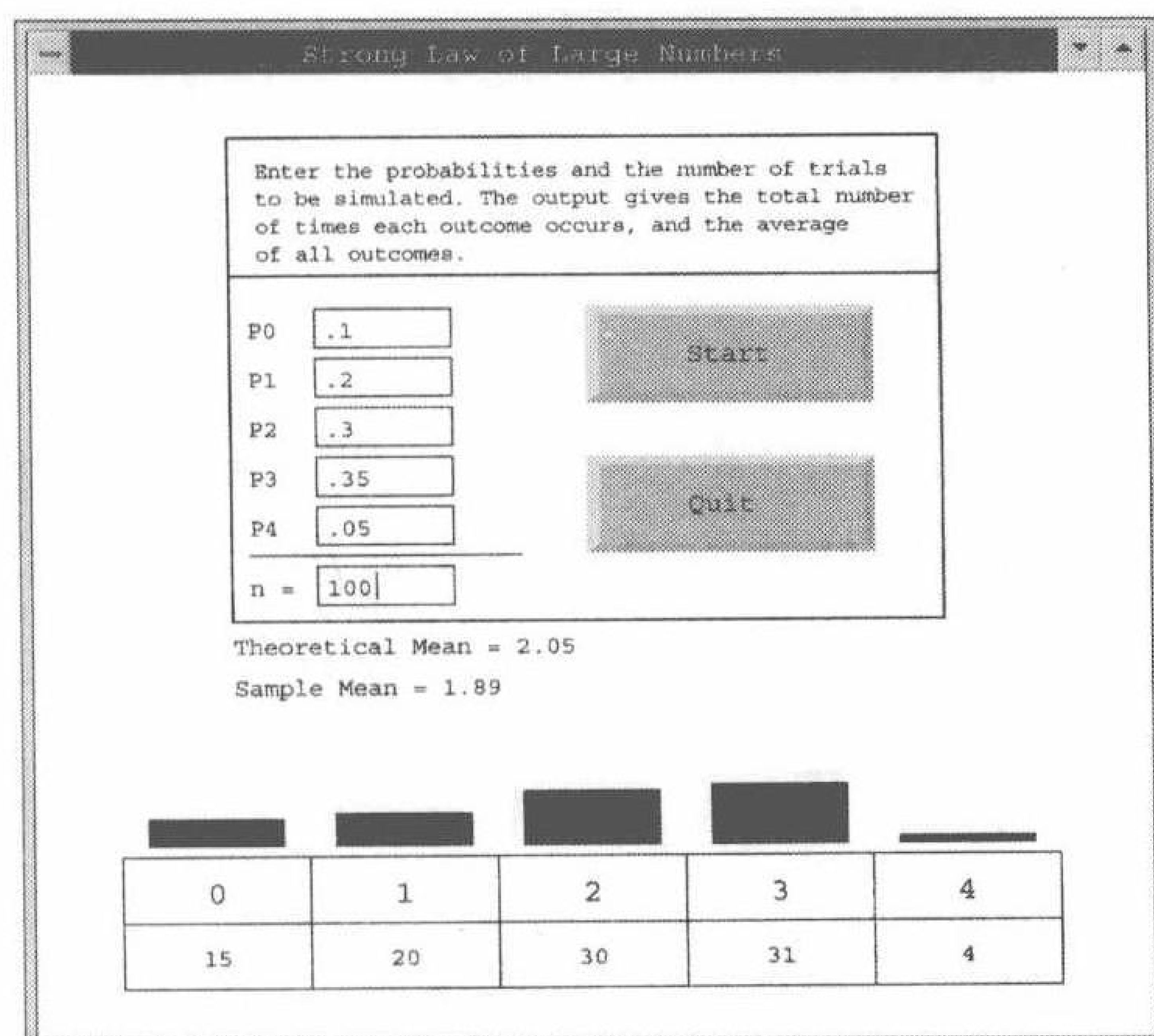


图 8.2(a) $n = 100$

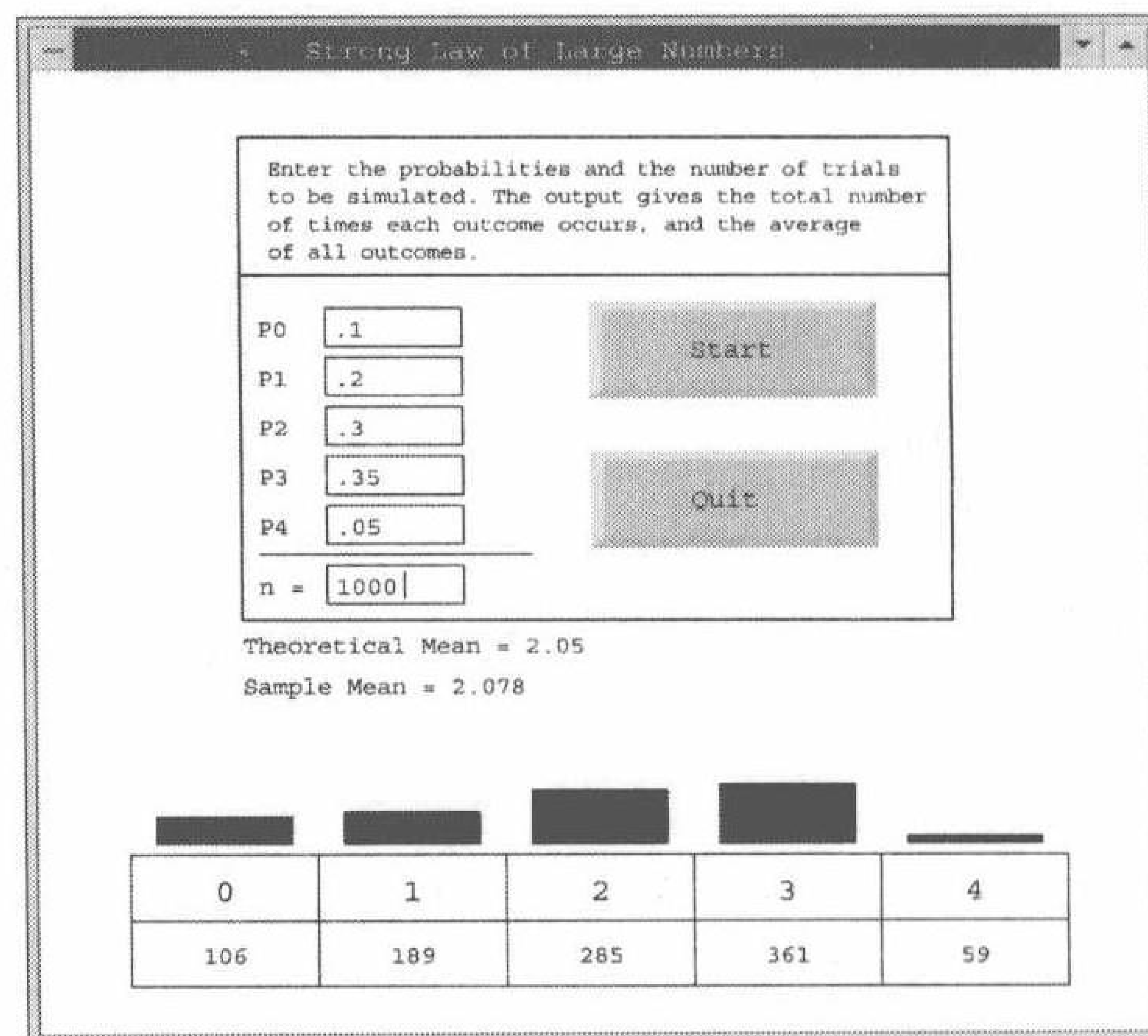


图 8.2(b) $n = 1000$

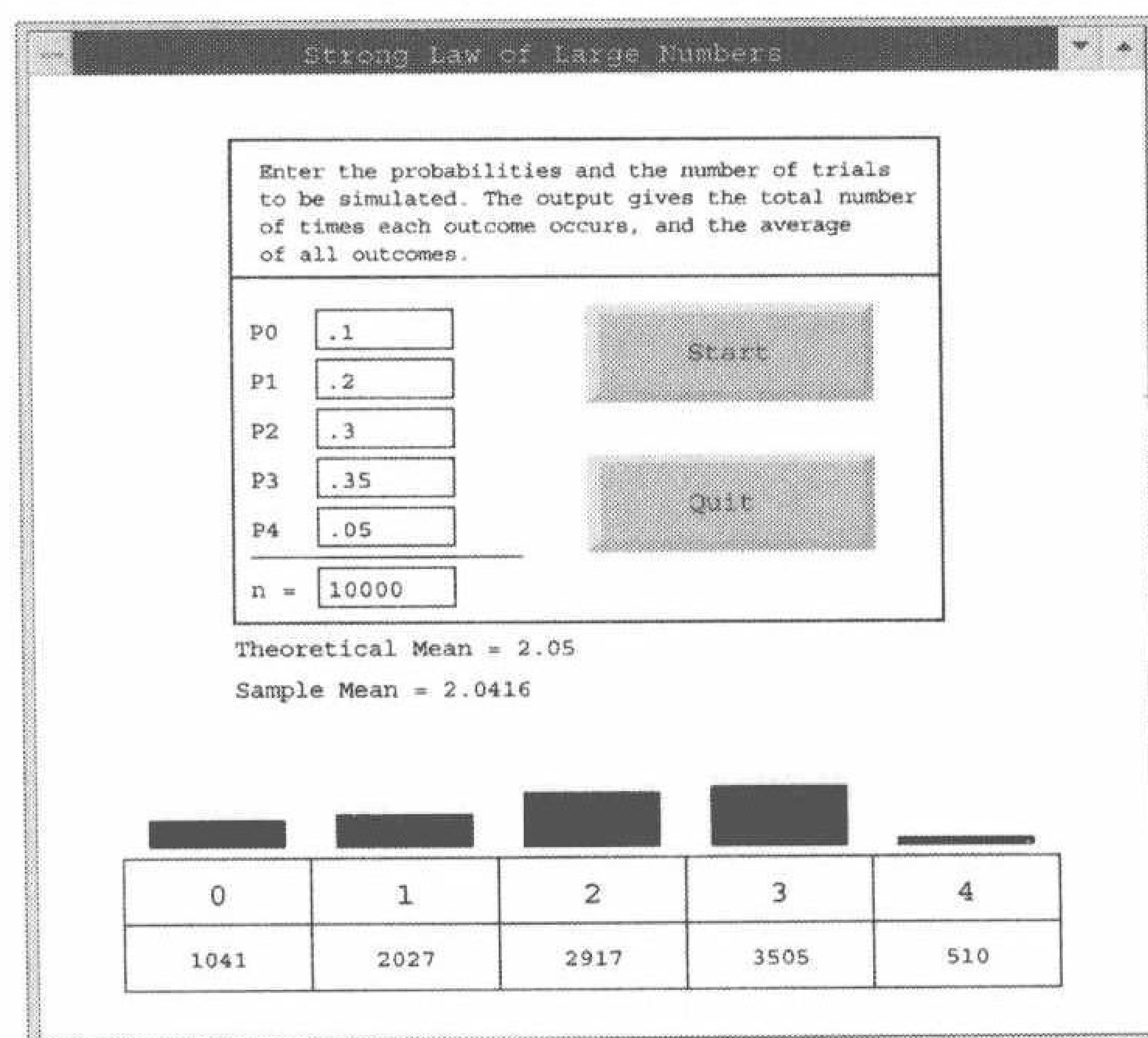


图 8.2(c) $n = 10\,000$

许多学生往往混淆弱大数律与强大数律. 弱大数律只能保证对充分大的 n^*, 随机变量 $(X_1+\cdots+X_{n^*})/n^*$ 靠近 μ. 但它不能保证对一切 $n > n^*$, $(X_1+\cdots+X_n)/n$ 也一定在 μ 的附近. 这样, $|(X_1+\cdots+X_n)/n-\mu|$ 可以无限多次离开 0(尽管出现较大偏离的频率不会很高.) 而强大数律能保证这种情况不会出现, 强大数律能够以概率为 1 地保证, 对于任何 $\varepsilon > 0$,

$$\left|\sum_{i=1}^{n}\frac{X_i}{n}-\mu\right|>\varepsilon$$

只可能出现有限次.

法国数学家博雷尔最早在伯努利随机变量的特殊情况下证明了强大数律. 而一般情况下的定理 4.1 的证明是由前苏联数学家 A. N. 科尔莫戈罗夫给出的.

8.5　其他不等式

有时候, 我们需要求出概率 $P\{X-\mu\geqslant a\}$ 的上界, 其中 a 为一个正数, 而均值 $\mu=E[X]$ 和方差 $\sigma^2=\mathrm{Var}(X)$ 是已知的. 由于当 $a>0$ 时, $X-\mu\geqslant a$ 蕴含 $|X-\mu|\geqslant a$, 利用切比雪夫不等式可知

$$P\{X-\mu\geqslant a\}\leqslant P\{|X-\mu|\geqslant a\}\leqslant\frac{\sigma^2}{a^2}\qquad a>0$$

然而, 下面的命题指出, 我们可以得到更好的上界 (上界越小, 就越好).

命题 5.1 (单边的切比雪夫不等式)　设 X 具有 0 均值和有限方差 σ^2, 则对任意 $a>0$,

$$P\{X\geqslant a\}\leqslant\frac{\sigma^2}{\sigma^2+a^2}$$

证明: 令 $b>0$, 注意到

$$X\geqslant a\Leftrightarrow X+b\geqslant a+b$$

故

$$P\{X\geqslant a\}=P\{X+b\geqslant a+b\}\leqslant P\{(X+b)^2\geqslant(a+b)^2\}$$

上式中, 由 $X+b\geqslant a+b>0$ 可推知 $(X+b)^2\geqslant(a+b)^2$, 故不等式成立. 再利用马尔可夫不等式

$$P\{X\geqslant a\}\leqslant P\{(X+b)^2\geqslant(a+b)^2\}\leqslant E[(X+b)^2]/(a+b)^2=(\sigma^2+b^2)/(a+b)^2$$

上式中, b 可以取任何常数, 取 $b=\sigma^2/a$, 便得到本命题的结论. 实际上, 当 $b=\sigma^2/a$ 时, $(\sigma^2+b^2)/(a+b)^2$ 达到极小值.　□

例 5a　设某工厂每周的产量是一个随机变量, 其期望为 $\mu=100$, 方差为 $\sigma^2=400$. 现在问这一周产量至少为 120 的概率的上界是什么?

解: 利用单边切比雪夫不等式

$$P\{X\geqslant 120\}=P\{X-100\geqslant 20\}\leqslant\frac{400}{400+20^2}=\frac{1}{2}$$

故本周产量至少为 120 的概率不会超过 1/2. 但如果直接利用马尔可夫不等式,

$$P\{X \geqslant 120\} \leqslant \frac{E[X]}{120} = \frac{5}{6}$$

这个上界就比较弱[①]. ■

现在设 X 具有均值 μ, 方差 σ^2, 由于 $X-\mu$ 与 $\mu-X$ 都具有均值 0 和方差 σ^2, 利用单边的切比雪夫不等式可知, 对于 $a>0$,

$$P\{X-\mu \geqslant a\} \leqslant \frac{\sigma^2}{\sigma^2+a^2}$$

$$P\{\mu-X \geqslant a\} \leqslant \frac{\sigma^2}{\sigma^2+a^2}$$

因此, 我们得到下面的推论.

推论 5.1 若 $E[X]=\mu, \mathrm{Var}(X)=\sigma^2$, 则对于 $a>0$, 下列不等式成立:

$$P\{X \geqslant \mu+a\} \leqslant \frac{\sigma^2}{\sigma^2+a^2}$$

$$P\{X \leqslant \mu-a\} \leqslant \frac{\sigma^2}{\sigma^2+a^2}$$

例 5b 有一个 200 人的集体, 由 100 个男人和 100 个女人组成, 将他们随机地分成 100 个组, 2 人一组, 求最多 30 个组由一男一女组成的概率之上界.

解: 将男人编号, 由 1 到 100, 对于 $i=1,\cdots,100$, 令

$$X_i = \begin{cases} 1 & \text{男人}i\text{所在的组内有女人} \\ 0 & \text{其他} \end{cases}$$

这样, 男女配成对的组数 X 可以表示为

$$X = \sum_{i=1}^{100} X_i$$

男人 i 与其他任何人配成一组的概率是相同的, 因此

$$E[X_i] = P\{X_i=1\} = \frac{100}{199}$$

对于 $i \neq j$, 由于男人 i 已经与一个女人配对的条件下, 男人 j 只可能跟 197 人配对, 其中 99 人为女人, 因此

$$P\{X_j=1|X_i=1\} = \frac{99}{197}.$$

① 上界越小, 结论越强, 若上界为 1, 这个结论就没有任何意义了. —— 译者注

$$E[X_iX_j] = P\{X_i = 1, X_j = 1\}$$
$$= P\{X_i = 1\}P\{X_j = 1|X_i = 1\} = \frac{100}{199} \times \frac{99}{197}$$

这样,

$$E[X] = \sum_{i=1}^{100} E[X_i] = 100 \times \frac{100}{199} \approx 50.25$$

$$\mathrm{Var}(X) = \sum_{i=1}^{100} \mathrm{Var}(X_i) + 2\sum_{i<j}\sum \mathrm{Cov}(X_i, X_j)$$
$$= 100 \times \frac{100}{199} \times \frac{99}{199} + 2 \times \binom{100}{2} \times \left[\frac{100}{199} \times \frac{99}{197} - \left(\frac{100}{199}\right)^2\right] \approx 25.126$$

由切比雪夫不等式

$$P\{X \leqslant 30\} \leqslant P\{|X - 50.25| \geqslant 20.25\} \leqslant \frac{25.126}{(20.25)^2} \approx 0.061$$

由此看出, 最多 30 对为一男一女的概率之上界为 6.1%. 但是, 也可以利用单边的切比雪夫不等式

$$P\{X \leqslant 30\} = P\{X \leqslant 50.25 - 20.25\} \leqslant \frac{25.126}{25.126 + (20.25)^2} \approx 0.058$$

这样, 将上界稍作改进. ■

当 X 的矩母函数已知时, 我们可以得到更加有效的 $P\{X \geqslant a\}$ 的上界, 令

$$M(t) = E[\mathrm{e}^{tX}]$$

为 X 的矩母函数. 对于 $t > 0$,

$$P\{X \geqslant a\} = P\{\mathrm{e}^{tX} \geqslant \mathrm{e}^{ta}\} \leqslant E[\mathrm{e}^{tX}]\mathrm{e}^{-ta} \qquad \text{利用马尔可夫不等式}$$

类似地, 对于 $t < 0$,

$$P\{X \leqslant a\} = P\{\mathrm{e}^{tX} \geqslant \mathrm{e}^{ta}\} \leqslant E[\mathrm{e}^{tX}]\mathrm{e}^{-ta}$$

这样, 我们得到了切尔诺夫界.

命题 5.2 (切尔诺夫界)

$$P\{X \geqslant a\} \leqslant \mathrm{e}^{-ta}M(t) \qquad \text{对一切} t > 0$$
$$P\{X \leqslant a\} \leqslant \mathrm{e}^{-ta}M(t) \qquad \text{对一切} t < 0$$

由于切尔诺夫界对一切适用的 t 都成立, 我们可以找到 t, 使 $\mathrm{e}^{-ta}M(t)$ 达到最小值. 例如, 为找到 $P\{X \geqslant a\}$ 的最小上界, 可求 $\mathrm{e}^{-ta}M(t)$ 在 $t > 0$ 上的最小值.

例 5c (标准正态分布的切尔诺夫界) 设 Z 是一个标准正态随机变量, 它的矩母函数为 $M(t) = \mathrm{e}^{t^2/2}$, $P\{Z \geqslant a\}$ 的切尔诺夫界为

$$P\{Z \geqslant a\} \leqslant \mathrm{e}^{-ta}\mathrm{e}^{t^2/2} \qquad \text{对一切 } t > 0$$

对于 t 在 $(0,\infty)$ 上变化, 当 $t=a$ 时 $\mathrm{e}^{-ta+t^2/2}$ 达到极小值. 这样, 对于 $a>0$,

$$P\{Z \geqslant a\} \leqslant \mathrm{e}^{-a^2/2}$$

同理, 对于 $a<0$,

$$P\{Z \leqslant a\} \leqslant \mathrm{e}^{-a^2/2}$$

■

例 5d (泊松随机变量的切尔诺夫界) 设 X 是一个泊松随机变量, 其参数为 λ, 其矩母函数为 $M(t)=\mathrm{e}^{\lambda(\mathrm{e}^t-1)}$, 因此, $P\{X \geqslant i\}$ 的切尔诺夫界为

$$P\{X \geqslant i\} \leqslant \mathrm{e}^{\lambda(\mathrm{e}^t-1)}\mathrm{e}^{-it} \qquad t>0$$

上式右边的极小化等价于 $\lambda(\mathrm{e}^t-1)-it$ 的极小化问题, 这个极小化问题当 $\mathrm{e}^t=i/\lambda$ 时达到极小值. 只有当 $i/\lambda>1$ 时, 相应的极小点 t 的值大于 0. 因此我们在 $i>\lambda$ 之假设下[①], 令 $\mathrm{e}^t=i/\lambda$, 可得

$$P\{X \geqslant i\} \leqslant \mathrm{e}^{\lambda(i/\lambda-1)} \times \left(\frac{\lambda}{i}\right)^i$$

即

$$P\{X \geqslant i\} \leqslant \frac{\mathrm{e}^{-\lambda}(e\lambda)^i}{i^i}$$

■

例 5e 设一个赌徒, 每次赌博输和赢的概率是相等的, 并且输赢与过去的历史是相互独立的, 每次输和赢的数目是一个单位, 设 X_i 表示第 i 次赌博赢的单位数. 则 X_i 相互独立, 且

$$P\{X_i=1\}=P\{X_i=-1\}=\frac{1}{2}$$

记 $S_n=\sum_{i=1}^n X_i$ 表示经过 n 次赌博后该赌徒的累计赢钱数, 我们求 $P\{S_n \geqslant a\}$ 的切尔诺夫界, 注意 X 的矩母函数为

$$E[\mathrm{e}^{tX}]=\frac{e^t+e^{-t}}{2}$$

利用 e^t 和 e^{-t} 的 McLaurin 展式, 得

$$\begin{aligned}
\mathrm{e}^t+\mathrm{e}^{-t} &= \left(1+t+\frac{t^2}{2!}+\frac{t^3}{3!}+\cdots\right)+\left(1-t+\frac{t^2}{2!}-\frac{t^3}{3!}+\cdots\right)\\
&=2\left\{1+\frac{t^2}{2!}+\frac{t^4}{4!}+\cdots\right\}\\
&=2\sum_{n=0}^{\infty}\frac{t^{2n}}{(2n)!} \leqslant 2\sum_{n=0}^{\infty}\frac{(t^2/2)^n}{n!} \qquad \text{由于}(2n)! \geqslant n!2^n\\
&=2\mathrm{e}^{t^2/2}
\end{aligned}$$

故

$$E[\mathrm{e}^{tX}] \geqslant \mathrm{e}^{t^2/2}$$

① 在 $i \leqslant \lambda$ 的情况下, 本例的上界可能不正确. ——译者注

利用独立随机变量和的矩母函数等于各随机变量矩母函数的乘积, 我们得到

$$E[e^{tS_n}] = (E[e^{tX}])^n \leqslant e^{nt^2/2}$$

再利用关于切尔诺夫界的公式可得

$$P\{S_n \geqslant a\} \leqslant e^{-ta}e^{nt^2/2} \qquad t > 0$$

通常为了求得更好的结果, 我们需要求上式右边的极小值, 或等价地, 求 $nt^2/2 - ta$ 的极小值. 利用二次式极小值的公式易知, 当 $t = a/n$ 时, $e^{-ta}e^{nt^2/2}$ 达极小值, 假定 $a > 0$, 此时 $t = a/n$ 取正值, 将这个值代入切尔诺夫的界中, 得

$$P\{S_n \geqslant a\} \leqslant e^{-a^2/2n} \qquad a > 0$$

例如, 由这个不等式, 可得

$$P\{S_{10} \geqslant 6\} \leqslant e^{-36/20} \approx 0.1653$$

经计算, 实际的概率值

$$\begin{aligned} P\{S_{10} \geqslant 6\} &= P\{\text{在10次赌博中至少赢8次}\} \\ &= \frac{\binom{10}{8} + \binom{10}{9} + \binom{10}{10}}{2^{10}} = \frac{56}{1024} \approx 0.0547 \end{aligned}$$

■

下面的不等式是关于期望的不等式.

定义 一个二次可微的实值函数 $f(x)$ 称为凸的, 若 $f''(x) \geqslant 0$ 对一切 x 成立; 类似地, 若 $f''(x) \leqslant 0$, 对一切 x 成立, 则称 f 为凹的.

凸函数的一些例子如 $f(x) = x^2, f(x) = e^{ax}, f(x) = -x^{1/n}, x \geqslant 0$. 若 $f(x)$ 为凸的, 则 $g(x) = -f(x)$ 就是凹的, 反之亦然.

命题 5.3 (詹生不等式) 设 $f(x)$ 为凸函数并且 $E[X]$ 存在且有限, 则

$$E[f(X)] \geqslant f(E[X])$$

证明: 将 $f(x)$ 在 $\mu = E[X]$ 处进行泰勒展开,

$$f(x) = f(\mu) + f'(\mu)(x - \mu) + \frac{f''(\xi)(x - \mu)^2}{2}$$

其中 ξ 是在 x 与 μ 之间的某个值. 由于 $f''(\xi) \geqslant 0$, 我们得到

$$f(x) \geqslant f(\mu) + f'(\mu)(x - \mu)$$

因此

$$f(X) \geqslant f(\mu) + f'(\mu)(X - \mu)$$

两边取期望得

$$E[f(X)] \geqslant f(\mu) + f'(\mu)E[X - \mu] = f(\mu)$$

命题得证. □

例 5f 一个投资者会遇到这样的问题, 她可以将所有钱财投入一个具有风险的计划, 其回报为一个随机变量, 期望为 m, 也可将钱财投入一个没有风险的计划, 这个计划的回报为 m, 且以概率为 1 地是一个常数, 两种方式的回报的期望都是 m, 她选什么为好? 现在假定她有一个效用函数 $u(R)$, 其中 R 为回报值, 且假定她期望获得最大的 $E[u(R)]$. 通常 $u(R)$ 为 R 的增函数, 第一种投资方式的效用为 $u(X)$, 第二种方式的功效为 $u(m)$, 若 $u(R)$ 为凹函数, 则由詹生不等式 $E[u(X)] \leqslant u(m)$, 她应选择没有风险的投资方式, 若 $u(R)$ 为凸函数, 则 $E[u(X)] \geqslant u(m)$, 她应选择具有风险的投资方式. ■

8.6 用泊松随机变量逼近独立的伯努利随机变量和的概率误差界

本节中, 我们要讨论用泊松随机变量逼近独立伯努利随机变量和的问题, 其中伯努利随机变量的均值为 $p_i, i=1,2,\cdots,n$.

开始时, 先设 $Y_1,\cdots,Y_n$ 为独立的泊松随机变量, 其参数为 $p_i, i=1,2,\cdots,n$, 现在我们指出, 对每一个 i, 由 Y_i 可构造出伯努利随机变量 X_i, 其参数为 p_i, 并满足条件

$$P\{X_i \neq Y_i\} \leqslant p_i^2$$

令 $U_1,\cdots,U_n$ 为独立的随机变量序列, 且与 $Y_1,\cdots,Y_n$ 独立. U_i 的分布由下式给出.

$$P\{U_i=1\}=1-(1-p_i)\mathrm{e}^{p_i} \quad P\{U_i=0\}=(1-p_i)\mathrm{e}^{p_i}$$

由于不等式 $\mathrm{e}^{-p} \geqslant 1-p$, 保证了上式中 $(1-p_i)\mathrm{e}^{p_i}$ 在 $(0,1)$ 区间内, 从而保证了定义 U_i 的合理性, 现在定义 X_i,

$$X_i=\begin{cases}0 & 若 Y_i=U_i=0\\ 1 & 其他\end{cases}$$

注意到

$$P\{X_i=0\}=P\{Y_i=0\}P\{U_i=0\}=\mathrm{e}^{-p_i}(1-p_i)\mathrm{e}^{p_i}=1-p_i$$
$$P\{X_i=1\}=1-P\{X_i=0\}=p_i$$

由 X_i 之定义看出, 若 $X_i=0$, 必有 $Y_i=0$. 故

$$\begin{aligned}&P\{X_i \neq Y_i\}=P\{X_i=1, Y_i \neq 1\}\\ =&P\{X_i=1, Y_i=0\}+P\{X_i=1, Y_i>1\}=P\{X_i=1, Y_i=0\}+P\{Y_i>1\}\\ =&P\{Y_i=0, U_i=1\}+P\{Y_i>1\}=\mathrm{e}^{-p_i}[1-(1-p_i)\mathrm{e}^{p_i}]+1-\mathrm{e}^{-p_i}-p_i\mathrm{e}^{-p_i}\\ =&p_i-p_i\mathrm{e}^{-p_i} \leqslant p_i^2 \qquad 利用 1-\mathrm{e}^{-p} \leqslant p\end{aligned}$$

由 X_i 之定义可知, X_i 只依赖于 U_i 和 Y_i, 而 $(Y_1,U_1),\cdots,(Y_n,U_n)$ 是相互独立的. 因此, $X_1,\cdots,X_n$ 也是相互独立的. 这样, 我们所构造的随机变量序列 $X_1,\cdots,X_n$ 是相互独立的伯努利随机变量序列, 其参数为 $p_1,\cdots,p_n$, 并且满足条件

$$P\{X_i \neq Y_i\} \leqslant p_i^2$$

记 $X=\sum_{i=1}^n X_i, Y=\sum_{i=1}^n Y_i$, 由上式可知

$$P\{X \neq Y\} \leqslant \sum_{i=1}^n p_i^2$$

现在设 A 为任何实数之集合, 则下列事件之集合恒等式成立

$$\{X \in A\}=\{X \in A, Y \in A\}+\{X \in A, Y \notin A\}$$

$$\{Y \in A\}=\{X \in A, Y \in A\}+\{Y \in A, X \notin A\}$$

由上式可知

$$P\{X \in A\}-P\{Y \in A\}=P\{X \in A, Y \notin A\}-P\{Y \in A, X \notin A\}$$

进一步可得

$$|P\{X \in A\}-P\{Y \in A\}| \leqslant P\{X \in A, Y \notin A\}+P\{Y \in A, X \notin A\}$$

又由于 $\{X \in A, Y \notin A\}$ 与 $\{Y \in A, X \notin A\}$ 互不相容, 它们之中任一事件发生, 必有 $X \neq Y$ 发生, 故

$$P\{X \in A, Y \notin A\}+P\{Y \in A, X \notin A\} \leqslant P\{X \neq Y\} \leqslant \sum_{i=1}^n p_i^2$$

综合起来, 可知, 对任何实数集合 A,

$$|P\{X \in A\}-P\{Y \in A\}| \leqslant \sum_{i=1}^n p_i^2$$

利用泊松随机变量的性质可知, $Y=\sum_{i=1}^n Y_i$ 也是泊松随机变量, 其参数为 $\lambda=\sum_{i=1}^n p_i$, 利用 Y 的分布列可得

$$\left|P\left\{\sum_{i=1}^n X_i \in A\right\}-\sum_{i \in A} \frac{\mathrm{e}^{-\lambda} \lambda^i}{i!}\right| \leqslant \sum_{i=1}^n p_i^2$$

注释 当所有的 p_i 都等于 p 时, X 就是二项随机变量, 上式变成

$$\left|\sum_{i \in A}\binom{n}{i} p^i(1-p)^{n-i}-\sum_{i \in A} \frac{\mathrm{e}^{-np}(np)^i}{i!}\right| \leqslant np^2$$ ■

小 结

马尔可夫和切尔诺夫不等式是概率论中十分重要的不等式, 可提供有关概率的上界. 马尔可夫不等式所涉及的是非负随机变量, 对于非负随机变量 X,

$$P\{X \geqslant a\} \leqslant \frac{E[X]}{a}$$

切比雪夫不等式是马尔可夫不等式的应用. 设 X 具有期望 μ, 方差 σ^2, 则对于每一个 $k>0$, 有

$$P\{|X-\mu| \geqslant k\sigma\} \leqslant \frac{1}{k^2}$$

此不等式中的随机变量 X 不再限于非负随机变量了.

概率论中两个最重要的理论成果是**中心极限定理**和**强大数律**, 它们都讨论独立同分布随机变量序列的性质. 若这些随机变量具有期望 μ 和方差 σ^2, 则中心极限定理说明当 n 充分大时, 这个序列的前 n 个变量的和的分布, 近似地为正态分布, 其期望为 $n\mu$, 方差为 $n\sigma^2$. 设 $\{X_i, i \geqslant 1\}$ 是这样的一个序列, 则对于每一个实数 a,

$$\lim_{n\to\infty} P\left\{\frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}} \leqslant a\right\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a} \mathrm{e}^{-x^2/2}\mathrm{d}x$$

强大数律只要求该序列中的随机变量具有有限均值 μ. 强大数律说明, 当 n 趋于无穷时, 这个序列的前 n 项的平均值以概率为 1 地趋于 μ. 由强大数律可知, 在独立重复试验中, 事件 A 出现的频率以概率为 1 地趋于概率 $P(A)$. 如果将以概率为 1, 解释成“一定”, 这样, 我们就验证了概率的频率定义的合理性.

习　题

1. 设 X 是一随机变量, 其期望和方差均为 20. 对于概率 $P\{0<X<40\}$ 有什么结论?

2. 根据过去的经验, 一个学生的期末成绩是一个随机变量, 其均值为 75.

(a) 给出学生超过 85 分的概率的一个上界.

假设还知道学生成绩的方差为 25.

(b) 对于学生成绩在 $65\sim85$ 之间的概率, 你有什么结论?

(c) 设参加考试学生人数为 n, n 多大时, 才能以 90% 以上的把握保证学生的平均成绩在 75 ± 5 这个范围内? 请不要利用中心极限定理.

3. 利用中心极限定理解习题 2(c).

4. 设 $X_1,\cdots,X_{20}$ 是独立泊松随机变量序列, 期望为 1.

(a) 利用马尔可夫不等式求 $P\left\{\sum_{i=1}^{20} X_i>15\right\}$ 的上界.

(b) 利用中心极限定理求 $P\left\{\sum_{i=1}^{20} X_i>15\right\}$ 的近似值.

5. 将 50 个数利用舍入法化成 50 个整数, 设舍入误差的分布是 $(-0.5,+0.5)$ 上的均匀分布. 求这 50 个整数的和与原来的和相差超过 3 的概率的近似值.

6. 连续地掷一枚骰子, 一直到点数总和超过 300 点为止, 求至少需掷 80 次的概率的近似值.

7. 一个人有 100 个灯泡, 每一个灯泡的寿命为指数分布, 其平均寿命为 5 小时. 他每次用一个灯泡, 灯泡灭了以后立即换上一个. 求 525 小时后, 他仍有灯泡可用的概率的近似值.

8. 在习题 7 中, 假定换一个灯泡需一定时间, 换灯泡时间为随机变量, 分布为 $(0, 0.5)$ 上的均匀分布. 问在 550 小时的时候, 所有灯泡都已经烧掉的概率 (近似值).

9. 设 X 为 Γ 随机变量, 其参数为 $(n,1)$. n 应该多大才能满足

$$P\left\{\left|\frac{X}{n}-1\right|>0.01\right\}<0.01?$$

10. 工程师认为一座桥的载荷 W 是一个正态随机变量, 其均值为 400, 标准差 40. 先假定车辆的重量为随机变量, 均值为 3, 标准差为 0.3. 约多少辆车在桥上时, 可使桥的结构遭破坏的概率超过 0.1? (单位: 1000 磅)

11. 许多人相信公司股票价格的每日涨跌幅是一个随机变量, 其期望为 0, 方差为 σ^2. 设 Y_n 为第 n 天的股票价格, 则

$$Y_n = Y_{n-1} + X_n, n \geqslant 1$$

其中, $X_1, X_2, \cdots$ 为独立同分布随机变量, 期望为 0, 方差为 σ^2. 现假定今日的股票价格为 100, 如果 $\sigma^2 = 1$, 对于 10 天以后股票价格超过 105 的概率, 你有什么结论?

12. 一共有 100 个元件, 有替换地使用, 即当元件 1 失效以后, 立刻换上元件 2, 元件 2 失效后, 立刻换上元件 3, 等等. 设元件 i 的寿命分布是指数分布, 其均值为 $10 + i/10, i = 1,2,\cdots,100$, 估计总寿命超过 1200 的概率. 如果这些寿命的分布为 $(0, 20 + i/5)$ 上的均匀分布, $i = 1,2,\cdots,100$, 重算上述概率.

13. 某教师给学生评分的均值为 74, 标准差为 14. 现该教师对两个班进行测验, 一个班有 25 人, 另一个班有 64 人.

(a) 估计在 25 人的班中平均成绩超过 80 分的概率 (近似值).

(b) 对于 64 人的班, 重复 (a) 的计算.

(c) 估计大班的平均成绩超过小班 2.2 分的概率.

(d) 估计小班的平均成绩超过大班 2.2 分的概率.

14. 设某元件对某电气系统是一个关键的部件, 当该元件失效后应该立即换上一个新的元件. 假定该元件的平均寿命为 100, 标准差 30(小时), 应该有多少备件, 才能以 0.95 以上的概率保证这个系统连续运行 2000 小时?

15. 一个保险公司有 10 000 个汽车投保人, 每个投保人索赔的期望值为 280 美元, 标准差为 800 美元. 求总索赔超过 2 700 000 美元的概率.

16. A. J. 需要处理 20 个文件, 这些文件的处理时间是独立同分布的随机变量序列, 均值为 50 分钟, 标准差为 10 分钟. M. J. 也有 20 个文件要处理, 他处理的时间也是独立同分布的随机变量序列, 不过均值为 52 分钟, 标准差为 15 分钟.

(a) 求 A. J. 在 900 分钟内完成任务的概率.

(b) 求 M. J. 在 900 分钟内完成任务的概率.

(c) 计算 A. J. 在 M. J. 之前完成任务的概率.

17. 在下列假定下重做例 5b: 男女成对的数目是近似正态的. 这个假定是否合理?

18. 在习题 2(a) 中进一步假定学生成绩的方差为 25, 重做该习题.

19. 一个湖中有 4 种鱼, 假定抓起来的鱼属于任一种的可能性都是相同的, 记 Y 为每一种鱼至少抓住一条时所需抓鱼的总条数.

(a) 找一个区间 (a,b), 使得 $P\{a \leqslant Y \leqslant b\} \geqslant 0.90$.

(b) 利用单边的切比雪夫不等式, 求出所需抓的鱼的条数使得至少以 90% 以上的把握保证 4 种鱼抓全.

20. 设 X 为非负随机变量, 其期望为 25. 关于下列的量, 我们会有什么结论?

(a) $E[X^3]$; (b) $E[\sqrt{X}]$; (c) $E[\ln X]$; (d) $E[\mathrm{e}^{-X}]$.

21. 设 X 为非负随机变量, 证明

$$E[X] \leqslant (E[X^2])^{1/2} \leqslant (E[X^3])^{1/3} \leqslant \cdots$$

22. 设在 例 5f 中投资者也可以将 $100\alpha\%(0<\alpha<1)$ 的资产投入一个有风险的计划, 而将 $100(1-\alpha)\%$ 的财产投入无风险的计划, 例 5f 的结论会有什么改变? [当她的资产是这样被分割时, 其投资回报是 $R=\alpha X+(1-\alpha)m$.]

23. 设 X 为泊松随机变量, 其均值为 20.

(a) 利用马尔可夫不等式求出 $p=P\{X \geqslant 26\}$ 的一个上界.

(b) 利用单边切比雪夫不等式求出 p 的一个上界.

(c) 利用切尔诺夫界得到 p 的一个上界.

(d) 利用中心极限定理求 p 的近似值.

(e) 利用程序求 p 的近似值.

理论习题

1. 设 X 具有方差 σ^2, 则 σ 称为 X 的标准差. 若 X 具有均值 μ 和标准差 σ, 指出

$$P\{|X-\mu| \geqslant k\sigma\} \leqslant \frac{1}{k^2}$$

2. 设 X 具有期望 μ 和标准差 σ. 则比值 $r \equiv |\mu|/\sigma$ 称为 X 的测量的信噪比. 其思想来源于 X 可写成 $\mu+(X-\mu)$, 其中 μ 为信号部分, $X-\mu$ 为噪音部分. 定义 $|(X-\mu)/\mu| \equiv D$ 为 X 与 μ 的相对偏差, 指出, 对于 $\alpha>0$

$$P\{D \leqslant \alpha\} \geqslant 1-\frac{1}{r^2\alpha^2}$$

3. 计算下列随机变量的信噪比, 即计算 $|\mu|/\sigma$, 其中 $\mu=E[X], \sigma^2=\mathrm{Var}(X)$.

(a) 泊松随机变量, 其均值为 λ. (b) 二项随机变量, 其参数为 (n,p). (c) 具有几何分布的随机变量, 其均值为 $1/p$.

(d) 在 (a,b) 上均匀随机变量. (e) 指数随机变量, 其均值为 $1/\lambda$. (f) 正态随机变量, 参数为 (μ,σ^2).

4. 设 $Z_n, n \geqslant 1$ 是一个随机变量序列, c 是一个常数, 使得对每一个 $\varepsilon>0$, 当 $n \to \infty$ 时, $P\{|Z_n-c|>\varepsilon\} \to 0$. 指出对任何有界连续函数 g, 有

$$E[g(Z_n)] \to g(c) \qquad n \to \infty$$

5. 设 $f(x)$ 为 $[0,1]$ 上定义的连续函数, 考虑函数

$$B_n(x)=\sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k} x^k(1-x)^{n-k}$$

(称为伯恩斯坦多项式) 证明

$$\lim_{n \to \infty} B_n(x)=f(x)$$

提示：设 $X_1, X_2, \cdots$ 为独立同分布的伯努利随机变量，其公共均值为 x，利用

$$B_n(x) = E\left[f\left(\frac{X_1 + \cdots + X_n}{n}\right)\right]$$

和理论习题 4，也可证明 $B_n(x)$ 一致地趋向于 $f(x)$，这就提供了分析中著名的魏尔斯特拉斯定理的概率证明，这个定理说明连续函数可用多项式一致逼近.

6. (a) 令 X 为离散随机变量，它的可能值为 $1, 2, \cdots$. 若 $P\{X = k\}$ 为非增序列，$k = 1, 2, \cdots$，证明

$$P\{X = k\} \leqslant 2\frac{E[X]}{k^2}$$

(b) 设 X 为非负随机变量且具有非增的密度函数，指出

$$f(x) \leqslant \frac{2}{x^2}E[X] \qquad 对一切\ x > 0$$

7. 掷一枚均匀的骰子共 100 次. 记 X_i 为第 i 次的结果，计算

$$P\left\{\prod_{i=1}^{100} X_i \leqslant a^{100}\right\} \qquad 1 < a < 6$$

的近似值.

8. 参数为 (t, λ) 的 Γ 随机变量，当 t 很大时近似于正态分布，解释这一结论.

9. 掷一枚均匀的硬币 1000 次. 若前 100 次都是正面朝上，你认为后面的 900 次正面朝上的比例是多少？

10. 设 X 是一个泊松随机变量，其均值为 λ. 指出对于 $i < \lambda$，

$$P\{X \leqslant i\} \leqslant \frac{\mathrm{e}^{-\lambda}(e\lambda)^i}{i^i}$$

11. 设 X 为二项随机变量，其参数为 (n, p). 指出，对于 $i > np$，

(a) $\mathrm{e}^{-ti}E[\mathrm{e}^{tX}]$ 当 t 满足 $\mathrm{e}^t = i(1-p)/[(n-i)p]$ 时达到最小值 (此处限定 $t > 0$).

(b) $P\{X \geqslant i\} \leqslant \frac{n^n}{i^i(n-i)^{n-i}}p^i(1-p)^{n-i}$.

12. 关于标准正态随机变量的切尔诺夫界为

$$P\{Z > a\} \leqslant \mathrm{e}^{-a^2/2} \qquad a > 0$$

指出，利用 Z 的密度函数的形式，可将上界压缩成

$$P\{Z > a\} \leqslant \frac{1}{2}\mathrm{e}^{-a^2/2} \qquad a > 0$$

13. 若随机变量 X 满足 $E[X] < 0$，同时存在 $\theta \neq 0$，使得 $E[\mathrm{e}^{\theta X}] = 1$，指出这个 θ 必须满足 $\theta > 0$.

自 检 习 题

1. 在某特许经销商那里，每周汽车销售量是一个随机变量，其期望值为 16，求下列事件的概率的上界：

(a) 下周销售量超过 18 辆；(b) 下周销售量超过 25 辆.

2. 假定自检习题 1 中汽车周销售量的方差为 9.

(a) 给出下周销售量在 11 到 22 之间 (含 11 和 22) 的概率的下界.

(b) 给出下周销售量超过 18 的概率的上界.

3. 设 $E[X]=75, E[Y]=75, \mathrm{Var}(X)=10, \mathrm{Var}(Y)=12, \mathrm{Cov}(X,Y)=-3$. 给出下列概率的上界:

(a) $P\{|X-Y|>15\}$; (b) $P\{X>Y+15\}$; (c) $P\{Y>X+15\}$.

4. 设工厂 A 每日生产量的期望为 20 个单位, 标准差为 3. 工厂 B 每日生产量的期望为 18 个单位, 标准差为 6. 假定两个厂的日产量是相互独立的, 求出今日 B 比 A 产量多的概率的上界.

5. 设某种元件的寿命 (单位: 天) 的密度函数为

$$f(x)=2x \qquad 0<x<1$$

当元件失效后, 立即用新的元件替换. 用 X_i 表示第 i 个元件的寿命, $S_n=\sum_{i=1}^n X_i$ 表示第 n 个元件的失效时刻, 失效率定义如下:

$$r=\lim_{n\to\infty}\frac{n}{S_n}$$

假设 $X_i, i\geqslant 1$ 相互独立, 求出 r.

6. 在自检习题 5 中, 为了以 90% 的把握维持运行 35 天, 需要准备多少元件?

7. 机修厂修理机器需 2 个阶段, 第一阶段所需时间为指数分布, 期望为 0.2 个小时. 第二阶段所需时间也是指数分布, 并且与第一阶段独立, 期望为 0.3 小时. 现在修理工有 20 台机器需要修理, 求出他在 8 小时内完成修理任务的概率的近似值.

8. 有一种赌博游戏, 每次赌博以 0.7 的概率输 1 元, 以 0.2 的概率输 2 元, 以 0.1 的概率赢 10 元. 求出该赌徒在前 100 次赌博后输的概率的近似值.

9. 在自检习题 7 中, 确定时间 t, 使得在时间 t 内完成 20 台机器的修理的概率约为 0.95.

10. 烟草公司宣称一枝香烟中尼古丁含量的均值为 2.2mg, 标准差为 0.3mg. 但实测 100 支香 烟中, 尼古丁含量的均值为 3.1mg, 现在假定烟草公司的说法是对的, 问 100 支香烟中尼古丁含量均值高于 3.1mg 的概率为多少?

11. 一共有 40 节电池, 每一节电池可能是类型 A 或类型 B, 是 A 或者 B 的概率是相等的 $\left(\frac{1}{2}\right)$. 类型 A 的寿命均值为 50, 标准差为 15, 类型 B 的寿命均值为 30, 标准差为 6.

(a) 求出 40 节电池的总寿命超过 1700 的概率 (近似值).

(b) 假定已知 20 节电池为 A 型, 20 节电池为 B 型, 现在求总寿命超过 1700 的概率的近似值.

12. 一诊所每天的志愿医生的人数是一个随机变量, 取值为 2,3,4. 并且这三种情况是等可能的. 不管当天来了多少志愿者, 每位志愿者所看的病人数为泊松随机变量, 期望为 30. 记 X 为当天由志愿者所看的病人数.

(a) 求 $E[X]$. (b) 求 $\mathrm{Var}(X)$. (c) 求 $P\{X>65\}$ 的近似值 (利用正态分布表).

13. 强大数律指出, 独立同分布的随机变量序列的前 n 项的算术平均以概率为 1 地收敛到他们的公共期望值. 现在的问题是: 前 n 项的几何平均收敛到什么? 即

$$\lim_{n\to\infty}\left(\prod_{i=1}^n X_i\right)^{1/n}=?$$

第 9 章　概率论的其他课题

9.1　泊松过程

在定义泊松过程以前, 我们把 h 的函数 $f(h)$ 称为 $o(h)$ 的定义叙述一下. f 称为 $o(h)$, 如果它满足

$$\lim_{h\to 0}\frac{f(h)}{h}=0.$$

f 为 $o(h)$ 是指当 $h\to 0$ 的过程中, f 的值比起 h 来, 还要小得多. 现在假定在时间轴上的一些随机点上发生了若干事件[①], 记 $N(t)$ 在时间区间 $[0,t]$ 上发生的事件的个数, 随机变量集合 $\{N(t),t\geqslant 0\}$ 称为具有速率 λ 的泊松过程 $(\lambda>0)$, 如果

(i) $N(0)=0$.

(ii) 在不相交的时间段内发生的事件数是相互独立的.

(iii) 在给定时间区间上发生的事件数的分布只跟时间区间的长度有关, 而与这个区间的位置无关.

(iv) $P\{N(h)=1\}=\lambda h+o(h)$.

(v) $P\{N(h)\geqslant 2\}=o(h)$.

条件 (i) 只说明这个过程在 $t=0$ 处开始, 并且 $N(0)=0$. 条件 (ii) 是独立增量假设, 例如, 在 t 以前发生的事件数 (即 $N(t)$) 与在区间 $(t,t+s)$ 内的事件数 (即 $N(t+s)-N(t)$) 是相互独立的随机变量. 条件 (iii) 是平稳增量假定, 即 $N(t+s)-N(t)$ 的分布与 t 无关.

在第 4 章, 基于泊松分布为二项分布的极限这一结论, 推导出 $N(t)$ 具有泊松分布, 其参数为 λt. 此处我们也要导出这一结果, 但方法稍有不同.

引理 1.1　对于速率为 λ 的 泊松过程,

$$P\{N(t)=0\}=\mathrm{e}^{-\lambda t}$$

证明: 记 $P_0(t)=P\{N(t)=0\}$, 我们由下式导出一个微分方程.

$$\begin{aligned}P_0(t+h)&=P\{N(t+h)=0\}=P\{N(t)=0,N(t+h)-N(t)=0\}\\&=P\{N(t)=0\}P\{N(t+h)-N(t)=0\}=P_0(t)[1-\lambda h+o(h)]\end{aligned}$$

---

① 此处的事件与前面讨论的具有概率的随机事件有所区别, 在“随机点上发生的事件”都是零概率事件, 有的作者称其为“随机点上发生的呼唤”. —— 译者注

上式中倒数第二个等式是独立性条件 (ii) 的结果, 而最后一个等式利用了 (iv) 和 (v). 因此

$$\frac{P_0(t+h)-P_0(t)}{h}=-\lambda P_0(t)+\frac{o(h)}{h}$$

再令 $h\to 0$, 得

$$P_0'(t)=-\lambda P_0(t)$$

或等价地

$$\frac{P_0'(t)}{P_0(t)}=-\lambda$$

将上式积分, 得

$$\ln P_0(t)=-\lambda t+c$$

或

$$P_0(t)=K\mathrm{e}^{-\lambda t}$$

再利用 $P_0(0)=P\{N(0)=0\}=1$, 我们得到

$$P_0(t)=\mathrm{e}^{-\lambda t}.$$

□

对于一个泊松过程, 用 T_1 表示第一个事件的发生时间, 对于 $n>1$, 记 T_n 为第 $n-1$ 个事件点到第 n 个事件点的时间间隔, 序列 $\{T_n, n=1,2\cdots\}$ 称为泊松过程的时间间隔序列. 例如, $T_1=5, T_2=10$ 表示泊松过程第一个事件发生在时刻 $t=5$, 第二个事件发生在时刻 $t=15$.

现在我们要确定 T_n 的分布. 首先我们指出, 随机事件 $\{T_1>t\}$ 发生的充要条件是泊松过程在 $[0,t]$ 内事件发生的个数为 0, 因此

$$P\{T_1>t\}=P\{N(t)=0\}=\mathrm{e}^{-\lambda t}$$

即 T_1 具有指数分布, 其均值为 $1/\lambda$. 现在计算

$$P\{T_2>t\}=E[P\{T_2>t|T_1\}]$$

但

$$\begin{aligned}P\{T_2>t|T_1=s\}&=P\{(s,s+t]\text{上有 0 个事件发生}|T_1=s\}\\&=P\{(s,s+t]\text{上有 0 个事件发生}\}=\mathrm{e}^{-\lambda t}\end{aligned}$$

上式中第二个等式是由独立性所得, 而第三个等式是由泊松过程的平稳性所得. 由上式可以得到两个结论, T_2 也具有指数分布, 其期望值为 $1/\lambda$. 同时 T_2 与 T_1 相互独立, 重复上述推论过程可得命题 1.1.

命题 1.1 速率为 λ 的泊松过程的时间间隔序列为独立同分布序列, 其公共分布为指数分布, 公共期望为 $1/\lambda$.

与泊松过程有关的另一个重要的量为 S_n, 第 n 个事件发生的时刻, 或者称第 n 个事件的到达时刻 (等待时间), 这个名词来源于服务系统的顾客到达时刻 (或系

统的等待时间). 易知,

$$S_n = \sum_{i=1}^{n} T_i \qquad n \geqslant 1$$

因此, 由 命题 1.1 和 第 5章6.1 节知, S_n 具有 Γ 分布, 其参数为 (n, λ). 即 S_n 的概率密度为

$$f_{S_n}(x) = \lambda \mathrm{e}^{-\lambda x} \frac{(\lambda x)^{n-1}}{(n-1)\,!} \qquad x \geqslant 0$$

现在我们可以证明 $N(t)$ 是一个泊松随机变量, 其均值为 λt.

定理 1.1 对于一个速率为 λ 的泊松过程,

$$P\{N(t) = n\} = \frac{\mathrm{e}^{-\lambda t}(\lambda t)^n}{n\,!}, \quad n = 0, 1, 2, \cdots$$

证明: 注意到第 n 个事件在 t 或 t 以前发生的充要条件是在 t 以前至少发生了 n 个事件, 即

$$N(t) \geqslant n \Leftrightarrow S_n \leqslant t$$

故

$$\begin{aligned} P\{N(t) = n\} &= P\{N(t) \geqslant n\} - P\{N(t) \geqslant n+1\} = P\{S_n \leqslant t\} - P\{S_{n+1} \leqslant t\} \\ &= \int_0^t \lambda \mathrm{e}^{-\lambda x} \frac{(\lambda x)^{n-1}}{(n-1)\,!} \mathrm{d}x - \int_0^t \lambda \mathrm{e}^{-\lambda x} \frac{(\lambda x)^n}{n\,!} \mathrm{d}x \end{aligned}$$

利用分部积分公式 $\int u\,\mathrm{d}v = uv - \int v\,\mathrm{d}u$, 其中 $u = \mathrm{e}^{-\lambda x}, \mathrm{d}v = \lambda[(\lambda x)^{n-1}/(n-1)!]\mathrm{d}x$ 得

$$\int_0^t \lambda \mathrm{e}^{-\lambda x} \frac{(\lambda x)^{n-1}}{(n-1)!} \mathrm{d}x = \mathrm{e}^{-\lambda t} \frac{(\lambda t)^n}{n!} + \int_0^t \lambda \mathrm{e}^{-\lambda x} \frac{(\lambda x)^n}{n!} \mathrm{d}x$$

由此可得定理之结论. □

9.2 马尔可夫链

考虑随机变量序列 $X_0, X_1, \cdots$, 假定这些随机变量的可能取值的集合为 $\{0, 1, \cdots, M\}$. 通常可将 X_n 解释为系统在时刻 n 的状态. 因此, 当 $X_n = i$ 时我们就说这个系统在时刻 n 处于状态 i. 随机变量序列称为马尔可夫链, 如果某时刻处于状态 i, 且存在固定的概率 P_{ij}, 使得下一时刻以概率 P_{ij} 处于状态 j. 即对所有 $i_0, \cdots, i_{n-1}, i, j$,

$$P\{X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \cdots, X_1 = i_1, X_0 = i_0\} = P_{ij}$$

值 $P_{ij}, 0 \leqslant i \leqslant M, 0 \leqslant j \leqslant M$ 称为马尔可夫链的转移概率. 它们满足 (为什么？)

$$P_{ij} \geqslant 0 \quad \sum_{j=0}^{M} P_{ij} = 1 \quad i = 0, 1, \cdots, M$$

可将 P_{ij} 排成如下一个方阵：

$$\begin{Vmatrix} P_{00} & P_{01} & \cdots & P_{0M} \\ P_{10} & P_{11} & \cdots & P_{1M} \\ \cdots & \cdots & \cdots & \cdots \\ P_{M0} & P_{M1} & \cdots & P_{MM} \end{Vmatrix}$$

这样的方阵称为转移概率矩阵.

有了转移概率阵和 X_0 的初始分布, 我们就可以计算所有有关的概率. 例如, $X_0,\cdots,X_n$ 的联合分布列可由下列方式得到

$$\begin{aligned} &P\{X_n=i_n,X_{n-1}=i_{n-1},\cdots,X_1=i_1,X_0=i_0\} \\ &\quad =P\{X_n=i_n|X_{n-1}=i_{n-1},\cdots,X_0=i_0\}P\{X_{n-1}=i_{n-1},\cdots,X_0=i_0\} \\ &\quad =P_{i_{n-1},i_n}P\{X_{n-1}=i_{n-1},\cdots,X_0=i_0\} \end{aligned}$$

继续这个步骤, 最后, 得到这个概率等于

$$P_{i_{n-1},i_n}P_{i_{n-2},i_{n-1}}\cdots P_{i_1,i_2}P_{i_0,i_1}P\{X_0=i_0\}.$$

例 2a 假设明天是否下雨只决定于今天是否下雨. 又假定, 若今天下雨, 则明天下雨的概率为 α, 若今天不下雨, 则明天下雨的概率为 β. 若用状态 0 表示下雨, 状态 1 表示不下雨. 这样, 天气系统成为一个两个状态的马尔可夫链, 其转移概率矩阵为

$$\begin{Vmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{Vmatrix}$$

即 $P_{00}=\alpha=1-P_{01},P_{10}=\beta=1-P_{11}$. ■

例 2b 一赌徒在赌博的时候, 每赌一局或者赢一个单位, 赢的概率为 p, 或输一个单位, 输的概率为 $1-p$. 当赌徒的赌本为 0 或 M 时, 则赌博停止. 此时, 赌徒的赌本是一个马尔可夫链, 其转移概率为

$$P_{i,i+1}=p=1-P_{i,i-1} \quad i=1,\cdots,M-1$$

$$P_{00}=P_{MM}=1$$

例 2c P. 埃伦费斯特和 T. 埃伦费斯特是两位物理学家 (又是夫妇), 他们提出了一个分子运动的理论模型. 设有两个坛子, 里面共有 M 个分子, 每一次随机地选择一个分子, 把它从原来的坛子移向另一个坛子, 记 X_n 表示第 1 个坛子经过 n 次转移以后的分子的个数. 则 $\{X_0,X_1,\cdots\}$ 是一个马尔可夫链, 其转移概率为

$$P_{i,i+1}=\frac{M-i}{M} \qquad 0\leqslant i\leqslant M$$

$$P_{i,i-1}=\frac{i}{M} \qquad 0\leqslant i\leqslant M$$

$$P_{ij}=0 \qquad 若|j-i|>1$$

■

对于马尔可夫链, P_{ij} 表示由状态 i 转入 j 的概率, 我们也可以定义两步转移的概率 $P_{ij}^{(2)}$, 它等于一个系统原来在状态 i, 经过两步转移以后到达状态 j 的概率. 即:

$$P_{ij}^{(2)} = P\{X_{m+2} = j | X_m = i\}$$

$P_{ij}^{(2)}$ 可以由 P_{ij} 经过下列方式计算得到

$$P_{ij}^{(2)} = P\{X_2 = j | X_0 = i\} = \sum_{k=0}^{M} P\{X_2 = j, X_1 = k | X_0 = i\}$$

$$= \sum_{k=0}^{M} P\{X_2 = j | X_1 = k, X_0 = i\} P\{X_1 = k | X_0 = i\} = \sum_{k=0}^{M} P_{kj} P_{ik}$$

一般情况下, 可求出 n 步转移概率 $P_{ij}^{(n)}$,

$$P_{ij}^{(n)} = P\{X_{n+m} = j | X_m = i\}$$

为计算 $P_{ij}^{(n)}$, 我们引入下列命题.

命题 2.1 (查普曼–科尔莫戈罗夫方程)

$$P_{ij}^{(n)} = \sum_{k=0}^{M} P_{ik}^{(r)} P_{kj}^{(n-r)} \qquad 0 < r < n$$

证明:

$$P_{ij}^{(n)} = P\{X_n = j | X_0 = i\} = \sum_{k} P\{X_n = j, X_r = k | X_0 = i\}$$

$$= \sum_{k} P\{X_n = j | X_r = k, X_0 = i\} P\{X_r = k | X_0 = i\} = \sum_{k} P_{kj}^{(n-r)} P_{ik}^{(r)} \qquad \square$$

例 2d (随机游动)　随机游动是一个可数无限状态空间中的马尔可夫链. 假定状态空间是 $\{0, \pm 1, \cdots\}$, 当质点处于状态 i 时, 它下一步会以 p 的概率往右移一步,以 $1-p$ 的概率往左一步. 这样, 质点的路径形成一个马尔可夫链, 其转移概率为

$$P_{i,i+1} = p = 1 - P_{i,i-1} \qquad i = 0, \pm 1, \cdots$$

现在设一个质点处于状态 i, 若它经过 n 步转移以后到达状态 j, 那么, 其中 $(n-i+j)/2$ 步是往右的, 而 $n-(n-i+j)/2 = (n+i-j)/2$ 步是往左的. 每一步往右的概率为 p, 并且独立于其他各步, 它恰好是一个二项概率,

$$P_{ij}^{n} = \binom{n}{(n-i+j)/2} p^{(n-i+j)/2} (1-p)^{(n+i-j)/2}$$

上式中二项系数 $\binom{n}{x}$ 当 x 不是小于或等于 n 的非负整数时, 定义为 0. 上述公式

可以重新写成

$$P_{i,i+2k}^{2n} = \binom{2n}{n+k} p^{n+k}(1-p)^{n-k} \qquad k = 0, \pm 1, \cdots, \pm n$$

$$P_{i,i+2k+1}^{2n+1} = \binom{2n+1}{n+k+1} p^{n+k+1}(1-p)^{n-k} \qquad k = 0, \pm 1, \cdots, \pm n, -(n+1) \qquad \blacksquare$$

虽然 $P_{ij}^{(n)}$ 是条件概率, 但是我们可以利用初始概率, 导出相应的无条件概率, 例如

$$P\{X_n = j\} = \sum_i P\{X_n = j | X_0 = i\} P\{X_0 = i\} = \sum_i P_{ij}^{(n)} P\{X_0 = i\}$$

对于很大一部分的马尔可夫链, $P_{ij}^{(n)}$ 收敛到一个数 π_j, 它不依赖于初始状态 i. 可以指出, 具有这种性质的一个充分条件是: 存在 n, $n > 0$, 使得

$$P_{ij}^{(n)} > 0 \qquad \text{对所有的} i, j = 0, 1, \cdots, M \tag{2.1}$$

满足公式 (2.1) 的马尔可夫链称为遍历的, 由 命题 2.1 可得

$$P_{ij}^{(n+1)} = \sum_{k=0}^{M} P_{ik}^{(n)} P_{kj}$$

上式中令 $n \to \infty$, 对于遍历的马尔可夫链, 可得

$$\pi_j = \sum_{k=0}^{M} \pi_k P_{kj} \tag{2.2}$$

利用恒等式 $1 = \sum_{j=0}^{M} P_{ij}^{(n)}$, 令 $n \to \infty$, 可得

$$\sum_{j=0}^{M} \pi_j = 1 \tag{2.3}$$

事实上, 可以证明 $\pi_j, 0 \leqslant j \leqslant M$ 是方程组 (2.2),(2.3) 的唯一解. 所有这些结论, 可综合成下面的一条定理, 但是此处没有给出证明.

定理 2.1 对于遍历的马尔可夫链,

$$\pi_j = \lim_{n \to \infty} P_{ij}^{(n)}$$

存在, 并且 $\pi_j, 0 \leqslant j \leqslant M$ 是下列方程组的唯一解:

$$\pi_j = \sum_{k=0}^{M} \pi_k P_{kj} \qquad \sum_{j=0}^{M} \pi_j = 1$$

例 2e 考虑例 2a, 我们假定如果今天下雨, 则明天下雨的概率为 α; 如果今天不下雨, 则明天下雨的概率为 β. 由定理 2.1, 下雨和不下雨的极限概率 π_0 和 π_1 由下面的方程组给出

$$\pi_0 = \alpha\pi_0 + \beta\pi_1 \quad \pi_1 = (1-\alpha)\pi_0 + (1-\beta)\pi_1 \quad \pi_0 + \pi_1 = 1$$

由这个方程组得到

$$\pi_0 = \frac{\beta}{1+\beta-\alpha} \quad \pi_1 = \frac{1-\alpha}{1+\beta-\alpha}$$

例如, $\alpha = 0.6, \beta = 0.3$, 则下雨的极限概率为 $\pi_0 = 3/7$. ■

当 n 很大时, 量 π_j 也等于马尔可夫链处于状态 j 所占的时刻的比例, $j = 0, 1, \cdots, M$. 直观上可以这样地看这一事实: 记 P_j 表示当 n 很大时, 链处于状态 j 的时刻所占的比例. (利用强大数定律可以证明, 对于遍历的马尔可夫链, 这个比例的极限存在, 并且是一常数.) 现在设链处于状态 k 的比例为 P_k, 由于由 k 转向 j 的概率为 P_{kj}. 这样, 这个链中, 由 k 转向 j 的比例为 $P_k \cdot P_{kj}$, 对所有状态 k 求和, $\sum_k P_k P_{kj}$ 就是链处于状态 j 的比例. 这样, 处于状态 j 的时刻的比例应满足

$$P_j = \sum_k P_k P_{kj}$$

同时下式是明显成立的

$$\sum_j P_j = 1$$

再利用定理 2.1, π_j 是方程 (2.2)(2.3) 的唯一解. 可推知 $P_j = \pi_j, j = 0, 1, \cdots, M$. π_j 处于状态 j 的时刻的比例的解释, 即使是非遍历的马尔可夫链也是正确的.

例 2f 现在我们再考虑例 2c, 假设我们感兴趣于恰好有 j 个分子在第一个盒子中的时刻的比例, $j = 0, 1, \cdots, M$, 根据刚才的讨论, 恰好有 j 个分子在第一个盒子就是马尔可夫链处于状态 j . 因此, 所问的问题就是马尔可夫链处于状态 j 的时刻的比例问题. 利用定理 2.1, 可化为求解方程组

$$\pi_0 = \pi_1 \times \frac{1}{M}$$

$$\pi_j = \pi_{j-1} \times \frac{M-j+1}{M} + \pi_{j+1} \times \frac{j+1}{M} \qquad j = 1, \cdots, M$$

$$\pi_M = \pi_{M-1} \times \frac{1}{M}$$

$$\sum_{j=0}^{M} \pi_j = 1$$

易知, 这组方程组的解为

$$\pi_j = \binom{M}{j}\left(\frac{1}{2}\right)^M \qquad j = 0, \cdots, M$$

这些就是马尔可夫链保持在状态 j 的时刻所占的比例. (习题 11 给出了一个如何估算前面问题的解的方法.) ■

9.3 惊奇、不确定性及熵

考虑在试验中事件 E 发生以后的情况. 当你知道一个事件 E 发生以后, 一定会有某种惊奇的感觉, 不过惊奇的程度可大可小. 这种惊奇是由于事件 E 发生所带来的信息所引起的, 而这些信息又与事件 E 的概率有关. 我们还是用例子来说明. 设某人掷骰子, 当人们听到两个骰子的总和是一个偶数时, 并不感到十分惊奇(该事件的概率为 1/2). 但是当他掷出总和为 12 时, 就感到惊奇了, 原因是“和为 12”这个事件出现的概率为 1/36.

本节中, 我们要将惊奇量化, 首先我们必须有这样一个共识, 当知道某事件发生以后, 感到惊奇的程度只跟这个事件的概率有关. 我们用 $S(p)$ 表示由概率为 p 的事件发生以后所产生的惊奇感觉的程度. 为了确定 $S(p)$ 的具体形式, 我们需要给出 $S(p)$ 应该满足的条件, 根据这些条件确定 $S(p)$ 的形式. 首先我们假定 $S(p)$ 对一切 $0 < p \leqslant 1$ 有定义, 但对于概率为 0 的事件没有定义.

关于惊奇的第一个条件是: 当听到一个必然事件发生时不会产生任何惊奇. 因此

公理 1 $S(1) = 0$.

第二个条件是越不容易发生的事件发生后, 造成的惊奇感觉就越大.

公理 2 $S(p)$ 是 p 的严格递减函数, 若 $p < q$, 则 $S(p) > S(q)$.

第三个条件是数学上的条件, p 的微小变动也会导致 $S(p)$ 的微小变动.

公理 3 $S(p)$ 是一个 p 的连续函数.

现在考虑两个独立事件 E, F, 假定 $P(E) = p, P(F) = q, P(EF) = pq$. 因此, 当听到 E, F 发生时, 相应的惊奇为 $S(pq)$. 现在假定 E 首先发生, 然后 F 发生. 而 $S(p)$ 表示听到 E 发生后的惊奇, 由此可知 $S(pq) - S(p)$ 表示听到 F 发生以后的附加的惊奇. 由于 E, F 相互独立, E 的发生并不影响 F 的概率, 因此, 这部分附加的惊奇应该是 $S(q)$. 这样, 我们有

公理 4 $S(pq) = S(p) + S(q) \qquad 0 < p \leqslant 1, 0 < q \leqslant 1$

现在我们要给出 $S(p)$ 的表达式.

定理 3.1 若 $S(\cdot)$ 满足公理 $1 \sim 4$, 则 $S(p) = -C\log_2 p$, 其中 C 为任意正数.

证明: 由公理 4, 知

$$S(p^2) = S(p) + S(p) = 2S(p)$$

由此, 利用归纳法得

$$S(p^m) = mS(p) \tag{3.1}$$

同时, 对任何正整数 n, $S(p) = S(p^{\frac{1}{n}} \cdots p^{\frac{1}{n}}) = nS(p^{\frac{1}{n}})$, 由此推得

$$S(p^{\frac{1}{n}}) = \frac{1}{n}S(p) \tag{3.2}$$

由式 (3.1), (3.2) 知

$$S(p^{m/n}) = mS(p^{\frac{1}{n}}) = \frac{m}{n}S(p).$$

或者

$$S(p^x) = xS(p) \tag{3.3}$$

其中 x 为正有理数, 但由于 S 为 p 的连续函数 (公理 3), (3.3) 式对于非负实数都成立. (证明此结论!).

现在, 对任意 $p, (0 < p \leqslant 1)$, 令 $x = -\log_2 p$, 或 $p = (1/2)^x$, 由 (3.3) 式得

$$S(p) = S\left(\left(\frac{1}{2}\right)^x\right) = xS\left(\frac{1}{2}\right) = -C\log_2 p$$

其中 $C = S(\frac{1}{2}) > S(1) = 0$. □

当 $C = 1$ 时, $S(p) = -\log_2(p)$, 其单位为**比特**($S(\frac{1}{2})$).

现在考虑一个随机变量, 它的可能取值为 $x_1, \cdots, x_n$, 相应概率为 $p_1, \cdots, p_n$. 当观察到 x_i 以后, 引起的惊奇为 $-\log_2(p_i)$, 由此可知, 当观察到随机变量 X 所引起的平均的惊奇为

$$H(X) = -\sum_{i=1}^{n} p_i \log_2 p_i$$

在信息论中 $H(X)$ 称为随机变量 X 的熵 (当 $p_i = 0$ 时, $0\log_2(0) = 0$). 可以证明, 当 p_i 相同时, $H(X)$ 达到其最大值.

$H(X)$ 表示得知 X 值以后所引起的平均惊奇程度, 但也可以认为 X 的不确定程度, 事实上, 在信息理论中, $H(X)$ 就是观测到 X 的值以后所接收的平均信息量, 因此, 惊奇程度, 不确定性和信息量是从不同角度来看待 X 的同一个特性.

现在考虑两个随机变量 X 和 Y , 它们分别取值于 $x_1, \cdots, x_n$ 和 $y_1, \cdots, y_n$, 其联合分布列为

$$p(x_i, y_j) = P\{X = x_i, Y = y_j\}$$

随机向量 (X,Y) 所含的不确定性 $H(X,Y)$ 为

$$H(X,Y) = -\sum_i \sum_j p(x_i, y_j) \log_2 p(x_i, y_j)$$

设 $Y = y_j$ 已观测到, 此时 X 在 $Y = y_j$ 的条件下的不确定性为

$$H_{Y=y_j}(X) = -\sum_i p(x_i|y_j) \log_2 p(x_i|y_j)$$

其中

$$p(x_i|y_j) = P\{X = x_i|Y = y_j\}$$

因此, 当 Y 被观测到以后, X 的平均不确定性为

$$H_Y(X) = \sum_j H_{Y=y_j}(X) p_Y(y_j)$$

其中 $p_Y(y_j) = P\{Y = y_j\}$.

命题 3.1 说明了 $H(X,Y)$ 与 $H(Y), H_Y(X)$ 之间的关系. X,Y 的不确定性等于 Y 的不确定性加上 Y 被观测到以后 X 的平均剩余不确定性.

命题 3.1

$$H(X,Y) = H(Y) + H_Y(X).$$

证明: 利用恒等式 $p(x_i, y_j) = p_Y(y_j) p(x_i|y_j)$,

$$\begin{aligned}
H(X,Y) &= -\sum_i \sum_j p(x_i, y_j) \log_2 p(x_i, y_j) \\
&= -\sum_i \sum_j p_Y(y_j) p(x_i|y_j) [\log_2 p_Y(y_j) + \log_2 p(x_i|y_j)] \\
&= -\sum_j p_Y(y_j) \log_2 p_Y(y_j) \sum_i p(x_i|y_j) \\
&\quad - \sum_j p_Y(y_j) \sum_i p(x_i|y_j) \log_2 p(x_i|y_j) \\
&= H(Y) + H_Y(X)
\end{aligned}$$

□

当 Y 被观测到以后, X 的不确定性应该减少, 这是信息论的一个基本结果. 为证明这个结论, 我们需要下面的引理 (证明作为习题):

引理 3.1

$$\ln x \leqslant x - 1 \qquad x > 0$$

等号只有 $x = 1$ 处成立.

定理 3.2

$$H_Y(X) \leqslant H(X)$$

上述等号成立的充要条件是 X, Y 相互独立.

证明:

$$
\begin{aligned}
H_Y(X)-H(X) &= -\sum_i\sum_j p(x_i|y_j)\log_2[p(x_i|y_j)]p(y_j)+\sum_i\sum_j p(x_i,y_j)\log_2 p(x_i)\\
&=\sum_i\sum_j p(x_i,y_j)\log_2\left[\frac{p(x_i)}{p(x_i|y_j)}\right]\\
&\leqslant \log_2 \mathrm{e}\sum_i\sum_j p(x_i,y_j)\left[\frac{p(x_i)}{p(x_i|y_j)}-1\right] \qquad \text{利用引理 3.1}\\
&=\log_2 \mathrm{e}\left[\sum_i\sum_j p(x_i)p(y_j)-\sum_i\sum_j p(x_i,y_j)\right]\\
&=\log_2 \mathrm{e}[1-1]=0
\end{aligned}
$$
□

9.4　编码定理及熵

设一离散随机变量 X 在 A 地被观测到, 然后通过一个通讯网络送到 B 处, 而通讯网络信号由 0 和 1 组成. 为了实现通讯的目的, 我们必须把 X 的可能值编成一个一个的 0 - 1 序列. 为了避免混乱, 要求编码后的序列, 不能出现一个序列是另一个序列的延长.

例如, X 可取 x_1,x_2,x_3,x_4 , 则一个可能的编码方式是

$$x_1\leftrightarrow 00 \quad x_2\leftrightarrow 01 \quad x_3\leftrightarrow 10 \quad x_4\leftrightarrow 11 \tag{4.1}$$

若 $X=x_1$, 则将 00 送到 B 处, 若 $X=x_2$, 则将 01 送到 B 处, 等等. 这就形成一个编码系统. 另一种可能的编码方式是

$$x_1\leftrightarrow 0 \quad x_2\leftrightarrow 10 \quad x_3\leftrightarrow 110 \quad x_4\leftrightarrow 111 \tag{4.2}$$

但是下面的编码是不容许的:

$$x_1\leftrightarrow 0 \quad x_2\leftrightarrow 1 \quad x_3\leftrightarrow 00 \quad x_4\leftrightarrow 01$$

这是因为 00 是 0 的延长, 01 也是 0 的延长.

编码理论中的一个任务是设计一个编码系统, 使得在传送过程中具有最小的期望码长. 例如, 若

$$P\{X=x_1\}=\frac{1}{2} \quad P\{X=x_2\}=\frac{1}{4} \quad P\{X=x_3\}=\frac{1}{8} \quad P\{X=x_4\}=\frac{1}{8}$$

若利用 (4.2) 传递, 则平均码长为 $\frac{1}{2}\cdot 1+\frac{1}{4}\cdot 2+\frac{1}{8}\cdot 3+\frac{1}{8}\cdot 3=1.75$, 若用 (4.1) 传递, 则平均码长为 2, 因此, 对于上面一组概率, (4.2) 比 (4.1) 更有效.

现在提出这样的问题, 对于给定的随机变量, 什么样的编码系统是最有效? 其结果是这样的, 对于任何编码系统, 其平均码长大于或等于 X 的熵, 这个结果就是

信息论中的无噪声编码定理, 为证明此结果, 我们需要下面的引理 4.1.

引理 4.1[①] 设 X 的取值范围为 $x_1,\cdots,x_N$. 为了把它们编成长度为 $n_1,\cdots,n_N$ 的二进序列 (不能让其中一个序列为另一个序列的延长), 充要条件为

$$\sum_{i=1}^{N}\left(\frac{1}{2}\right)^{n_i}\leqslant 1$$

证明: 对于正整数 $n_1,\cdots,n_N$, 记 w_j 表示 n_i 中等于 j 的个数, $j=1,2,\cdots$. 为了使得它们形成编码系统, 显然, $w_1\leqslant 2$, 又由于不容许一个码为另一个码的延长对于 w_2, 必须满足 $w_2\leqslant 2^2-2w_1$, 其中 2^2 是码长为 2 的所有二进序列个数, 而 $2w_1$ 就是将长度为 1 的序列延长成 2 位序列的个数. 一般情况下据相同的理由, w_n 应满足

$$w_n\leqslant 2^n-w_12^{n-1}-w_22^{n-2}-\cdots-w_{n-1}2\quad n=1,2,\cdots \tag{4.3}$$

事实上, 仔细一想, 若有一组 $n_i, i=1,2,\cdots,N$, 满足上述条件, 就可能找到相应的二进序列将 x_i 进行编码, 并且 x_i 相应的码长为 n_i, 因此, (4.3) 是将 $x_1,\cdots,x_N$ 编成码长为 $n_1,\cdots,n_N$ 的编码系统的充要条件.

将 (4.3) 改写成

$$w_n+w_{n-1}2+\cdots+w_12^{n-1}\leqslant 2^n\quad n=1,2,\cdots$$

两边除以 2^n, 充要条件变成

$$\sum_{j=1}^{n}w_j\left(\frac{1}{2}\right)^{j}\leqslant 1\quad \text{对一切}n\text{成立} \tag{4.4}$$

由于 n 为任意的, 容易看出, 这个充要条件变成

$$\sum_{j=1}^{\infty}w_j\left(\frac{1}{2}\right)^{j}\leqslant 1$$

由于 w_j 是 $n_1,\cdots,n_N$ 中等于 j 的个数, 于是

$$\sum_{j=1}^{\infty}w_j\left(\frac{1}{2}\right)^{j}=\sum_{i=1}^{N}\left(\frac{1}{2}\right)^{n_i}$$ □

现在给出定理 4.1

定理 4.1 (无噪声编码定理) 设 X 取值于 $\{x_1,\cdots,x_N\}$, 其相应的概率为 $p(x_1),\cdots,p(x_N)$. 设有一个编码系统, 将 x_i 编成 n_i 位的二进序列, 则

$$\sum_{i=1}^{N}n_ip(x_i)\geqslant H(X)=-\sum_{i=1}^{N}p(x_i)\log_2 p(x_i)$$

① 在本章末, 译者给出该引理的另一种证明. —— 编者注

证明： 记 $P_i=p(x_i), q_i=2^{-n_i}/\sum_{j=1}^{N}2^{-n_j}, i=1,\cdots,N$. 关于这两组数, 我们有

$$-\sum_{i=1}^{N}P_i\log_2\left(\frac{P_i}{q_i}\right)=-\log_2\mathrm{e}\sum_{i=1}^{N}P_i\ln\left(\frac{P_i}{q_i}\right)=\log_2\mathrm{e}\sum_{i=1}^{N}P_i\ln\left(\frac{q_i}{P_i}\right)$$

$$\leqslant\log_2\mathrm{e}\sum_{i=1}^{N}P_i\left(\frac{q_i}{P_i}-1\right)\qquad\text{利用引理 3.1}$$

$$=0\qquad\text{由于}\sum_{i=1}^{N}P_i=\sum_{i=1}^{N}q_i=1$$

由此可得

$$-\sum_{i=1}^{N}P_i\log_2 P_i\leqslant-\sum_{i=1}^{N}P_i\log_2 q_i=\sum_{i=1}^{N}n_iP_i+\log_2\left(\sum_{j=1}^{N}2^{-n_j}\right)$$

$$\leqslant\sum_{i=1}^{N}n_iP_i\qquad\text{利用引理4.1}$$

□

这个不等式即定理之结论.

例 4a　考虑随机变量 X, 其分布列为

$$p(x_1)=\frac{1}{2}\quad p(x_2)=\frac{1}{4}\quad p(x_3)=p(x_4)=\frac{1}{8}$$

由于

$$H(X)=-\left[\frac{1}{2}\log_2\frac{1}{2}+\frac{1}{4}\log_2\frac{1}{4}+\frac{1}{4}\log_2\frac{1}{8}\right]=\frac{1}{2}+\frac{2}{4}+\frac{3}{4}=1.75$$

现在考虑编码

$$x_1\leftrightarrow 0\quad x_2\leftrightarrow 10\quad x_3\leftrightarrow 110\quad x_4\leftrightarrow 111\tag{4.5}$$

对于这组编码, 平均码长为 $\sum_{n_i}n_ip(x_i)=1.75=H(X)$. 由定理 4.1 知, 不会再有比这一组编码更有效的编码了. ■

对于大部分随机变量来说, 不会存在一组编码系统, 使得平均码长达到下界 $H(X)$. 但是可以存在一个编码系统, 使得平均码长与 $H(X)$ 之间的误差小于 1. 为此, 记 n_i 为满足下列条件的整数

$$-\log_2 p(x_i)\leqslant n_i<-\log_2 p(x_i)+1$$

由于

$$\sum_{i=1}^{N}2^{-n_i}\leqslant\sum_{i=1}^{N}2^{\log_2 p(x_i)}=\sum_{i=1}^{N}p(x_i)=1$$

利用引理 4.1, 我们能够构造一组编码 (0−1 序列), 使得其长度为 $n_i(i=1,\cdots,N)$, n_i 对应于 x_i. 此时, 这组编码的平均长度为 $L=\sum_{i=1}^{N}n_ip(x_i)$. 显然 L 满足

$$-\sum_{i=1}^{N}p(x_i)\log_2 p(x_i)\leqslant L<-\sum_{i=1}^{N}p(x_i)\log_2 p(x_i)+1$$

或

$$H(X) \leqslant L < H(X) + 1$$

例 4b 现独立抛掷 10 次硬币, 每次正面朝上的概率为 p, 现在要把这个信息由 A 端送 B 端. 试验的结果为 $X = (X_1, X_2, \cdots, X_{10})$, 其中

$$X_i = \begin{cases} 1 & 第\ i\ 次抛掷正面朝上 \\ 0 & 第\ i\ 次抛掷反面朝上 \end{cases}$$

根据刚才得到的结果, 必定存在至少一个编码系统, 具平均码长 L 满足

$$H(X) \leqslant L \leqslant H(X) + 1.$$

由于 X_i 为相互独立的随机变量, 依命题 3.1 和定理 3.2, 得

$$H(X) = H(X_1, \cdots, X_n) = \sum_{i=1}^{N} H(X_i) = -10[p \log_2 p + (1-p)\log_2(1-p)]$$

设 $p = 1/2$, 则 $H(X) = 10$, 此时, 利用 $X = x$ 作为编码系统, 其平均码长为 10. 因此, 不会存在比 $X = x$ 本身这个编码系统更有效的编码系统. 例如, 前 5 次掷硬币, 得到正面朝上, 后 5 次反面朝上, 可将 1111100000 直接进行传送.

然而, 当 $p \neq 1/2$, 我们可以找到一组编码, 使得平均码长比 10 小. 例如 $p = 1/4$, 此时

$$H(X) = -10\left(\frac{1}{4}\log_2\frac{1}{4} + \frac{3}{4}\log_2\frac{3}{4}\right) = 8.11$$

我们可以找到一组编码, 其平均码长小于 9.11.

一个简单做法如下, 将 $(X_1, \cdots, X_{10})$ 分成 5 对, 2 个随机变量形成一对, 编码方法如下:

$$\begin{array}{llllll} X_i = \mathrm{T}, X_{i+1} = \mathrm{T} & \leftrightarrow & 0 & X_i = \mathrm{T}, X_{i+1} = \mathrm{H} & \leftrightarrow & 10 \\ X_i = \mathrm{H}, X_{i+1} = \mathrm{T} & \leftrightarrow & 110 & X_i = \mathrm{H}, X_{i+1} = \mathrm{H} & \leftrightarrow & 111 \end{array}$$

此处, $i = 1, 3, 5, 7, 9$, H 表示正面朝上, T 表示反面朝上. 这样可把 10 次掷硬币结果通过一对一对的编码将信号传送出去.

例如, 试验结果为 TTTHHTTTTH, 此时编码为 010110010, 其平均码长为

$$5\left[1\left(\frac{3}{4}\right)^2 + 2\left(\frac{1}{4}\right)\left(\frac{3}{4}\right) + 3\left(\frac{1}{4}\right)\left(\frac{3}{4}\right) + 3\left(\frac{1}{4}\right)^2\right] = \frac{135}{16} \approx 8.44 \qquad \blacksquare$$

到此为止, 我们讨论的传送都是无噪声传送, 在 A 端送出一个信号, 在 B 端接收到的是与 A 端完全相同的信号. 由于随机干扰, 在实际通讯中, 往往产生误差, 例如发送端发送的消息为 00101101, 而接收收端变成 01101101.

现在设在发送端发出一位 (0 或 1), 在接收端将以概率 p 收到正确的信号, 并且各位之间的传送是相互独立的, 这样的通讯系统称为二进对称通道. 现在设通道的参数 $p = 0.8$, 并假设传送的信号由很多位组成. 由于每位有 0.2 的概率误传, 若不

经过技术处理而直接传送, 这种错误是很严重的. 一种减少误传信号的方法是将信号重复 3 次, 然后在译码过程中按多数原则, 即按照下表中方法进行编码和译码.

编　码	解　码	编　码	解　码
$0 \to 000$	$\left.\begin{matrix}000\\001\\010\\100\end{matrix}\right\}\to 0$	$1 \to 111$	$\left.\begin{matrix}111\\110\\101\\011\end{matrix}\right\}\to 1$

这种编码的方法, 如果传输过程中最多只有一个错误. 那么, 通过译码, 还能得到正确的信号, 因此, 传送一位错误的概率变成

$$(0.2)^3+3(0.2)^2(0.8)=0.104$$

这是一个很大的改善, 事实上, 只要在编码时, 重复足够多次, 就可以将误传概率变成任意小. 例如下面的系统.

编　码	解　码
$0 \to 17$个0 $1 \to 17$个1	多数原则

可将传送一位的误差概率缩成小于 0.01.

上述编码方法的问题是, 虽然把传送误差减少了, 但是在通讯过程中, 每一位传送的信号也相应减少了 (见表 9.1).

表 9.1　重复传输的编码系统

每一位的误差概率	传送中每一位实际传送的信号
0.20	1
0.10	0.33(= 1/3)
0.01	0.06(= 1/17)

看起来, 想要减少误差概率, 不可避免地也需降低传送的效率. 然而, 这个结论是不正确的, 在信息论中有香农建立的*噪声编码定理*(noisy coding theorem), 其结果是惊人的, 现在我们叙述这个定理.

> **定理 4.2** (噪声编码定理)　存在一个数 C, 使得对于任何 $R<C$, 以及任何 $\varepsilon>0$, 存在一个编码译码系统, 使得传送一位的误差概率小于 ε, 而每一位实际传送的信号为 R, 最大的 C 的值 C^* 称为通道容量, 对于二进对称系统,
>
> $$C^*=1+p\log_2 p+(1-p)\log_2(1-p)$$

小　结

具有速率 λ 的泊松过程是一族随机变量 $\{N(t), t\geqslant 0\}$, 它涉及时间轴上的随机

点上发生的事件, 例如, $N(t)$ 表示在时间区间 $(0,t]$ 上发生事件的个数. 泊松过程由下面几个条件所界定.

(i) 在不相重的时间区间上发生的事件数是相互独立的.

(ii) 在一个区间内发生事件的个数的分布只依赖于这个区间的长度.

(iii) 在一个时刻只发生一个事件.

(iv) 事件发生的速率为 λ.

可以证明 $N(t)$ 是泊松随机变量, 其期望值为 λt. 此外, 两个相邻的事件之间的时间间隔 $T_i(i \geqslant 1)$ 是独立同分布的指数随机变量, 参数为 λ.

取值于 $\{0,1,\cdots,M\}$ 的随机变量序列 $\{X_n, n \geqslant 0\}$ 称为具有转移概率 P_{ij} 的马尔可夫链, 如果对于一切 $n, i_0, \cdots, i_n, i, j$,

$$P\{X_{n+1}=j|X_n=i, X_{n-1}=i_{n-1}, \cdots, X_0=i_0\}=P_{ij}$$

如果我们将 X_n 解释为时刻 n 的状态, 然后马尔可夫链可以解释为一串顺序的状态序列, 如果某一时刻在状态 i, 下一时刻在状态 j 的概率为 P_{ij}, 它与以前的历史相互独立. 有许多马尔可夫链, 处于状态 j 的概率当 $n \to \infty$ 时有一个极限概率, 它不依赖于初始状态, 若用 $\pi_j(j=0,\cdots,M)$ 表示这些概率, 它们是下列方程组的唯一解:

$$\pi_j = \sum_{i=0}^{M} \pi_i P_{ij} \qquad j=0,1,\cdots,M$$

$$\sum_{j=1}^{M} \pi_j = 1$$

进一步, π_j 也是马尔可夫链当 n 充分大时, 处于状态 j 的时刻的比例.

设 X 是取值于 $\{x_1,\cdots,x_n\}$ 的随机变量, 其相应的概率为 $\{p_1,\cdots,p_n\}$, 量

$$H(X) = -\sum_{i=1}^{n} p_i \log_2(p_i)$$

称为随机变量 X 的熵, 它可以解释为 X 具有不确定性的大小, 也可以解释为当观测到 X 以后所接受的平均信息量. 熵在二进制编码理论中有很重要的应用.

理论习题

1. 顾客按照速率为 λ 的泊松过程的方式到达一个银行, 设在前一个小时内有两位顾客到达, 求下列事件的概率:

   (a) 都在前 20 分钟内到达; (b) 在前 20 分钟内至少有一个顾客已经到达.

2. 假设在高速公路的某处每分钟通过的汽车数是一个泊松随机变量, 速率 $\lambda=3$ 辆 / 分. 现在设某人不顾一切地要冲过该公路, 他走过公路的时间为 s 秒. (假定他在公路上时, 若刚好有一辆车通过这一路口, 则这个人一定会受伤.) 求他不会受伤的概率. 考虑 $s=2,5,10,20$ 的情况, 求出其概率.

3. 假定在习题 2 中这个人比较机警, 在穿越公路这段时间内 (时间为 s 秒), 若只有一辆汽车通过, 他不会受伤. 但是, 当这段时间内有 2 辆以上的车经过时, 他一定会受伤. 求他穿越公路时不会受伤的概率, 并计算 $s = 5, 10, 20, 30$(秒) 时的概率值.

4. 设一共有 3 个白球和 3 个黑球, 将它们放在 2 个坛子里, 每个坛子里放 3 个球. 若第一个坛子里含 i 个白球, 则这个系统处于状态 $i, i = 0, 1, 2, 3$. 每一步从两个坛子中各随机地拿出一个球, 将第一个坛子中的球放回到第二个坛子, 将第二个坛子中的球放回到第一个坛子. 记 X_n 为 n 步以后的状态, 计算马尔可夫链 $\{X_n, n \geqslant 0\}$ 的转移概率.

5. 在例 2a 中, 假定今天下雨的概率为 0.5, 计算第 4 天下雨的概率 ($\alpha = 0.7, \beta = 0.3$).

6. 计算理论习题 4 中的极限概率.

7. 称一个转移概率矩阵为双重随机的, 如果它还满足

$$\sum_{i=0}^{M} P_{ij} = 1 \qquad j = 0, 1, 2, \cdots, M$$

如果这个马尔可夫链是遍历的, 证明 $\Pi_j = \dfrac{1}{M+1}, j = 0, 1, \cdots, M$.

8. 某人的精神可能处于 3 种状态, 兴奋 (c), 平静 (s), 郁闷 (g), 下面是精神状态的转移概率矩阵:

	c	s	g
c	0.7	0.2	0.1
s	0.4	0.3	0.3
g	0.2	0.4	0.4

这个矩阵是这样解释的: 以 s 行为例, 这一行表示若今天他比较平静, 那么他明天处于兴奋、平静和郁闷的概率分别为 0.4, 0.3, 0.3. 其余各行的解释是类似的. 求这个人处于兴奋天数所占的比例.

9. 假定明天是否下雨只依赖于过去两天的气候状况, 特别地, 若昨天和今天都下雨, 那么明天下雨的概率是 0.8; 若昨天下雨, 今天不下雨, 则明天下雨的概率为 0.3; 如果昨天不下雨, 今天下雨, 则明天下雨的概率为 0.4; 如果昨天和今天都不下雨, 则明天下雨的概率为 0.2. 求下雨天的比例.

10. 一个人每天出去跑步, 他出去的时候可从前门出去, 也可从后门出去, 前门出或后门出是等可能的. 他回家的时候也是等可能地从前门或后门回来. 他一共有 5 双运动鞋, 放在两个门的门边, 出去的时候随便从门边穿一双出去, 回来的时候就把鞋脱在门边. 他出去的时候如果门边没有鞋, 就只好光脚跑步.

(a) 建立一个马尔可夫链, 给出状态空间和转移概率阵.

(b) 求他光脚跑步次数所占的比例.

11. 在例 2f 中,

(a) 验证 Π_j 满足方程组.

(b) 对于给定的分子, 求出它在第一个盒子里的 (极限) 概率.

(c) 当时间很长以后, 事件“第 j 个分子落在第一个盒子中”, $j \geqslant 1$, 是否独立?

(d) 解释为什么 (b) 中极限概率是这样的.

12. 设某人掷两枚均匀骰子, 并计算所得点数之和, 求这个和数的熵.

13. 设 X 取 n 个值, 其相应的概率为 $P_1, \cdots, P_n$, 证明当 $P_i = 1/n, i = 1, 2, \cdots, n$ 时, $H(X)$ 达到极大值, 计算此时 $H(X)$ 的值.

14. 某人连续掷两枚均匀骰子, 令

$$X = \begin{cases} 1 & \text{若其点数之和等于6} \\ 0, & \text{其他} \end{cases}$$

令 Y 为第一次掷骰子所得到的点数, 计算 (a) $H(Y)$, (b) $H_Y(X)$, (c) $H(X, Y)$.

15. 一枚硬币, 抛掷时正面朝上的概率为 $p = 2/3$, 现连续抛掷 6 次, 计算试验结果的熵.

16. 一个随机变量可取 n 个值 $x_1, \cdots, x_n$, 相应的概率值为 $p(x_i), i = 1, 2, \cdots, n$. 我们希望问一系列问题, 每次只回答"是"或"否", 例如"是否 $X = x_1$?"或"是否 X 取 x_1, x_2 或 x_3 之一个?", 等等, 为了得到 X 的值, 你对平均问问题的次数有什么结论?

17. 对任何离散随机变量 X 和函数 f, 推出 $H(f(X)) \leqslant H(X)$.

18. 设 X 是在 A 端发送的 0 − 1 信号, Y 为 B 端接收的信号, $H(X) - H_Y(X)$ 称为传输率. 作为 $P\{X = 1\} = 1 - P\{X = 0\}$ 的函数, 当传输率达到最大时, 这个值称为传输通道的容量. 现设通道是一个二进对称通道, 即通道满足 $P\{Y = 1|X = 1\} = P\{Y = 0|X = 0\} = p$. 证明, 当 $P\{X = 1\} = 1/2$ 时, 传输率达到最大值 $1 + p\log_2 p + (1 - p)\log_2(1 - p)$(这个最大值就是通道的容量).

自 检 习 题

1. 一个泊松过程, 平均每小时发生 3 个事件.
   (a) 在早晨 8:00~10:00 之间没有事件发生的概率是多少?
   (b) 在早晨 8:00~10:00 之间发生的事件数的期望值是多少?
   (c) 在下午 2:00 以后, 第 5 个事件发生的期望时间是多少?

2. 某一个零售店的客流情况是这样的: 顾客按泊松过程规律到达商店, 速率为 λ(人 / 每小时). 已知在开门的第一个小时内有两个顾客到达, 求下列事件的概率:
   (a) 两个人都在前 20 分钟内到达; (b) 至少有一个人在 30 分钟以前到达.

3. 在路上, 每 5 辆卡车中有 4 辆后面跟一辆小汽车, 每 6 辆小汽车后面跟一辆卡车. 问在路上卡车所占比例有多大?

4. 某镇的天气分成晴、雨、阴. 如果当天下雨, 那么第二天要么是晴天, 要么是阴天, 两者具有相同的可能性. 如果当天不下雨, 那么以 1/3 的概率在下一天会维持原状. 如果第二天气候改变的话, 它会等可能地变到另外两种状态之一. 从长期来看, 晴天的比例有多大? 雨天的比例有多大?

5. 设 X 可取 5 个值, 其相应概率为 $0.35, 0.2, 0.2, 0.2, 0.05$. Y 也可取 5 个值, 其概率分别为 0.05,0.35,0.1,0.15,0.35.
   (a) 指出 $H(X) > H(Y)$; (b) 利用理论习题 13, 给出 $H(X) > H(Y)$ 的直观解释.

参考文献

9.1 节和 9.2 节

[1] KEMENY,J., L. SNELL,and A. KNAPP. *Denumerable Markov Chains.* New York: D.Van Nostrand Company, 1966.

[2] PARZEN,E. *Stochastic Processes.* San Francisco: Holden-Day, Inc., 1962.

[3] ROSS, S. M. *Introduction to Probability Models.* 9th ed. San Diego, Calif.: Academic Press, Inc., 2007.①

[4] ROSS, S. M. *Stochastic Processes.* 2nd ed. New York: John Wiley & Sons, Inc., 1996.

9.3 节和 9.4 节

[5] ABRAMSON, N. *Information Theory and Coding.* New York: McGraw-Hill Book Company, 1963.

[6] MCELIECE, R. *Theory of Information and Coding.* Reading, Mass.: Addison-Wesley Publishing Co., Inc., 1977.

[7] PETERSON, W., and E. WELDON. *Error Correcting Codes.* 2nd ed. Cambridge,Mass.: The MIT Press, 1972.

引理 4.1 设 X 的取值范围为 $x_1, \cdots, x_N$. 为了把它们编成长度为 $n_1, \cdots, n_N$ 的 0−1 序列 (不能让其中一个序列为另一个序列的延长), 充要条件为

$$\sum_{i=1}^{N}\left(\frac{1}{2}\right)^{n_i} \leqslant 1$$

证明: 为了证明引理, 我们再强调一些概念: 编码系统就是某些长度可不一样的 $0-1$ 序列的有限集合, 这个集合必须满足两个条件: 集合的各元素必须是互不相同的 $0-1$ 序列, 并且任意一个元素 (即 $0-1$ 序列) 都不是另一元素的延长. 编码系统的元素 (即 $0-1$ 序列) 称为码, 以区别于系统外的 $0-1$ 序列. 编码系统的码的个数 N 称为系统的大小. 每一个码 (即 $0-1$ 序列) 的长度称为码长. 编码系统的各码的码长形成一个集合 $\{n_1, \cdots, n_N\}$, 对于码长的集合, 可定义 $w_i = \{n_1, \cdots, n_N\}$ 中 $n_j = i$ 的个数, 即 w_i 为编码系统中码长为 i 的码的个数. 现在先证明, 对于一个编码系统, 其相应的 w_n 满足

$$w_n \leqslant 2^n - w_1 2^{n-1} - \cdots - w_{n-1}2, \quad n = 1, 2, \cdots \tag{4.3}$$

事实上, $w_1 \leqslant 2(n=1)$ 是十分明显的, 因为长度为 1 的 $0-1$ 序列一共有 2 个. 故编码系统中码长为 1 的个数不超过 2. 对于 $n=2$, 长度为 2 的 $0-1$ 序列的个数为 2^2, 其中有 $w_1 \times 2$ 个

① 本书中文版和英文影印版均已由人民邮电出版社出版. —— 编者注

为码长为 1 的码的延长, 这些长度为 2 的 $0-1$ 序列就不能作为码了. 因此, $w_2 \leqslant 2^2 - w_1 \times 2$, 即 (4.3) 对 $n=2$ 成立. 如此继续下去, 可得 (4.3) 对一切 n 成立.

将 (4.3) 改写成

$$w_n + w_{n-1}2 + \cdots + w_1 2^{n-1} \leqslant 2^n \qquad n = 1, 2, \cdots$$

两边除以 2^n, (4.3) 就变成

$$\sum_{j=1}^{n} w_j \left(\frac{1}{2}\right)^j \leqslant 1 \qquad \text{对一切 } n \text{ 成立} \tag{4.4}$$

由于 n 为任意的, 容易看出, 这个条件变成

$$\sum_{j=1}^{\infty} w_j \left(\frac{1}{2}\right)^j \leqslant 1$$

由于 w_j 是 $n_1, \cdots, n_N$ 中等于 j 的个数

$$\sum_{i=1}^{N} \left(\frac{1}{2}\right)^{n_i} = \sum_{j=1}^{\infty} w_j \left(\frac{1}{2}\right)^j \leqslant 1$$

现在设 X 的取值范围为 $x_1, \cdots, x_N$. 若把它编成码长为 $n_1, \cdots, n_N$ 的编码系统, 即 $\{x_1, \cdots, x_N\}$ 与编码系统中的码建立了一一对应关系. 根据刚才的讨论知, 关系式

$$\sum_{i=1}^{N} \left(\frac{1}{2}\right)^{n_i} \leqslant 1$$

必定成立. 这就是引理的必要性.

现在证明充分性. 设 $\{n_1, \cdots, n_N\}$ 为一个正整数序列, 满足条件

$$\sum_{i=1}^{N} \left(\frac{1}{2}\right)^{n_i} \leqslant 1$$

我们指出必定存在一个大小为 N 的编码系统, 使得相应的码长之集合为 $\{n_1, \cdots, n_N\}$. 我们还采用必要性证明中的记号 $w_i, i \geqslant 1$. 这样, 充分性条件中的不等式变成

$$\sum_{j=1}^{\infty} w_j \left(\frac{1}{2}\right)^j \leqslant 1.$$

由上式可推知 $w_1 \leqslant 2$, 可取 w_1 个长度为 1 的码. 在不等式中取 2 项, 可得 $w_2 \leqslant 2^2 - 2w_1$, 这说明在长度为 2 的 $0-1$ 序列中, 最多有 $2^2 - 2w_1$ 个 $0-1$ 序列, 这些长度为 2 的 $0-1$ 序列和原来的长度为 1 的码放在一起形成一个编码系统. 由于 $w_2 \leqslant 2^2 - 2w_1$, 显然可以选择 w_2 个长度为 2 的 $0-1$ 序列使得一共 $w_1 + w_2$ 个 $0-1$ 序列成为一个大小为 $w_1 + w_2$ 的编码系统. 依次类推, 可以选出 $w_1 + w_2 + w_3 + \cdots$ 个 $0-1$ 序列, 形成一个编码系统. 将这些码的长度列出来就是集合 $\{n_1, n_2, \cdots, n_N\}$. 有了这个编码系统, 就可以建立 $\{x_1, \cdots, x_N\}$ 与编码系统的一一对应关系而完成编码任务. □

第 10 章 模　　拟

10.1 引　　言

怎样确定单人纸牌游戏的胜出概率? 所谓单人纸牌游戏就是用某种固定的方法玩 52 张牌, 一副牌的顺序确定以后, 玩牌人的胜负就完全确定. 一种合理的假设是一副牌的 (52)! 种可能的顺序是等可能的, 然后数一数有多少种顺序会胜出, 最后计算胜出的概率. 然而这种方法显然不现实, 因为 (52)! 种顺序是相当大的数量. 而且即使一副牌的顺序已知, 也只有按规则玩牌以后才能知道玩牌人是否胜出.

看起来, 确定一副纸牌的胜出的概率是数学的难题. 然而, 在应用科学中, 试验是非常有价值的技术, 对于单人纸牌游戏, 试验就是玩一次牌, 或者可以编制一个计算计程序, 让机器去玩牌. 经过几次玩牌以后, 比如 n 次, 令

$$X_i = \begin{cases} 1 & i\text{次玩牌胜出} \\ 0 & \text{其他} \end{cases}$$

此时 $X_i, i = 1, 2, \cdots, n$ 是独立的伯努利随机变量,

$$E[X_i] = P\{\text{第}i\text{次玩牌胜出}\}$$

由强大数定律

$$\sum_{i=1}^{n} \frac{X_i}{n} = \frac{\text{赢的次数}}{\text{玩牌总次数}}$$

以概率为 1 地收敛到 $P\{\text{单人纸牌游戏中玩家赢}\}$. 玩大量次数的纸牌游戏以后, 赢牌的频率的就作为赢的概率的估计值. 用试验的方法来确定概率值的方法称为模拟.

为了使计算机实现模拟, 我们必须产生 $(0,1)$ 上均匀分布的随机变量的值, 这些值称为随机数. 大部分计算机有一内置程序, 称为随机数发生器, 它产生一个伪随机数序列. 这个伪随机数序列无法与 $(0,1)$ 均匀随机数序列区别开来. 大部分随机数是这样产生的, 它有一个初始值 X_0, 称为种子, 以及正数 a, c, m, 然后令

$$X_{n+1} = (aX_n + c) \text{ modulo } m \qquad n \geqslant 0$$

上述记号表示将 $aX_n + c$ 除以 m, 取其余数, 得到下一个数 X_{n+1}. 这样, 每个 X_n 是 $0, 1, \cdots, m-1$ 之一, X_n/m 近似地在 $(0,1)$ 上均匀分布. 可以指出, 适当地选择 a, c, m, 可以产生一个序列, 它就好像 $(0,1)$ 区间上的均匀随机变量序列.

开始模拟时, 我们假定能够模拟 $(0,1)$ 上均匀分布的随机变量, 并用随机数序列这一术语表示 $(0,1)$ 上均匀分布的随机变量的一组样本.

在单人纸牌游戏的例子中, 我们先给出一个 52 张牌的一个顺序, 然后让计算机去玩牌, 然而顺序的抽取必须是 $(52)!$ 种顺序中等可能地取出. 因此, 我们必须产生一个随机的排列. 下面的程序指出怎样利用随机数产生一个随机排列, 其方法如下, 先从 n 个对象中随机地选定一个对象，将这个选定的对象放在 n 这个位置上, 再在剩下 $n-1$ 个对象中随机选出一个对象放在 $(n-1)$ 这个位置上, 最后所有的对象都放在相应的位置上, 一个随机排列就产生了. 怎样从若干个对象中随机地选定一个对象? 其方法是将这些对象排成一个任意的顺序, 再利用随机数确定这个顺序中的一个位置, 在这个位置上的对象就是随机地确定的对象.

例 1a(产生一随机排列) 假设我们要将对象 $1,2,\cdots,n$ 做成一个排列, 使得所有 $n!$ 种排列都是等可能地成为选中的排列. 从任一初始排列开始, 经过 $n-1$ 步, 可产生一个随机排列, 而每一步将其中两个对象交换一下位置, 在整个过程中, 我们用 $X(i)$ 表示在位置 i 上的对象. 其算法如下.

1. 考虑一个初始排列, $X(i)$ 表示在位置 i 上的对象. [例如, 令 $X(i)=i, i=1,\cdots,n$.]

2. 产生一个随机变量 N_n, N_n 是数集 $\{1,2,\cdots,n\}$ 上的均匀随机变量.

3. 将 $X(N_n)$ 与 $X(n)$ 交换位置, 交换以后, $X(n)$ 就是原来的 $X(N_n)$, 并且将这个对象固定在位置 n. [例如, $n=4$, 初始状态 $X(i)=i, i=1,2,3,4$, 若 $N_4=3$, 此时, 新的排列成为 $X(1)=1, X(2)=2, X(3)=4, X(4)=3$. 而 3 这个对象此后不改位置, 永远放在位置 4 上.]

4. 产生随机数 N_{n-1}, 它均匀地分布在整数集 $\{1,2,\cdots,n-1\}$ 上.

5. 交换 $X(N_{n-1})$ 与 $X(n-1)$ 的位置. [若 $N_3=1$, 新的排列变成 $X(1)=4, X(2)=2, X(3)=1, X(4)=3$.]

6. 产生 N_{n-2} , 它是取值于 $\{1,2,\cdots,n-2\}$ 的均匀随机变量.

7. 交换 $X(N_{n-2})$ 和 $X(n=2)$. [若 $N_2=1$, 此时新的排列成为 $X(1)=2, X(2)=4, X(3)=1, X(4)=3$.]

8. 产生 $N_{n-3},\cdots$, 直到 N_2 产生, 然后将 $X(N_2)$ 与 $X(2)$ 交换位置, 得到最后的排列.

执行这个算法, 必须产生 $\{1,2,\cdots,k\}$ 上等可能地取值的随机变量, 其方法如下, 取一个随机数 U, 即 U 的分布为 $(0,1)$ 上均匀分布. 注意 kU 为 $(0,k)$ 上均匀分布. 因此

$$P\{i-1<kU<i\}=\frac{1}{k} \quad i=1,\cdots,k$$

取 $N_k=[kU]+1$, 其中记号 $[x]$ 表示 x 的整数部分, 即不超过 x 的最大整数, 则 N_k 就会在 $\{1,\cdots,k\}$ 上均匀分布. 最后, 我们可把算法简明地写成下列几条:

第 1 步 令 $X(1),\cdots,X(n)$ 为 $1,2,\cdots,n$ 的任意排列. [例如, $X(i)=i,i=1,\cdots,n$.]

第 2 步 令 $I=n$.

第 3 步 产生一个随机数 U, 令 $N=[IU]+1$.

第 4 步 交换 $X(N)$ 与 $X(I)$ 的位置.

第 5 步 将 I 的值减去 1, 如果 $I>1$, 则转向第 3 步.

第 6 步 $X(1),\cdots,X(n)$ 就是随机排列.

上述产生的随机排列是十分有用的. 例如, 一个统计工作者希望比较 m 种不同的处理的效应, 试验对象共 n 个, 他决定将 n 个对象分成 m 个不同的组, 每一个组含 n_i 个对象, 显然 $\sum_{i=1}^{m}n_i=n$. 对于第 i 个组施以处理 i, 为了清除任何偏向, (例如, 若把最好的个体放在同一组, 就会将处理的效果和个体的"好坏"作用相混淆, 造成偏差.) 我们必须将各个个体随机地分入各组. 怎样做才能完成这个任务?[①]

一个简单而有效的方法是将对象编号, 由 1 号到 n 号, 同时, 产生一个 $1,2,\cdots,n$ 的随机排列 $X(1),\cdots,X(n)$, 将编号为 $X(1),\cdots,X(n_1)$ 的对象归入第一组, 将编号 $X(n_1+1),\cdots,X(n_1+n_2)$ 编为第二组, 一般地, 将编号为 $X(n_1+\cdots+n_{j-1}+k),k=1,\cdots,n_j$ 的对象编成第 j 组, 然后对于第 j 组施行第 j 种处理. ■

10.2 具有连续分布函数的随机变量的模拟技术

本节中, 我们提供两种模拟随机变量的一般方法, 而这些随机变量具有连续分布函数.

10.2.1 反变换方法

基于下列命题, 我们可用反变换方法模拟具有连续分布函数的随机变量.

命题 2.1 设 U 为 $(0,1)$ 上均匀随机变量, F 为任意一个连续分布函数, 定义随机变量

$$Y=F^{-1}(U)$$

则 Y 具有分布函数 F. [$F^{-1}(x)$ 是方程 $F(y)=x$ 的解.]

证明:

$$F_Y(a)=P\{Y\leqslant a\}=P\{F^{-1}(U)\leqslant a\} \tag{2.1}$$

① 对于 $m=2$, 第 6 章例 2g 提供了另一种随机分组方法. 本处介绍的方法比较快, 但需要更多的存储空间.

由于 $F(x)$ 是一个单调函数, $F^{-1}(U) \leqslant a$ 成立的充要条件是 $U \leqslant F(a)$. 因此, 由 (2.1) 式可得

$$F_Y(a) = P\{U \leqslant F(a)\} = F(a) \qquad \square$$

由命题 2.1, 我们首先产生一个随机数 U, 然后令 $X = F^{-1}(U)$, 可知 X 具有连续分布函数 $F(x)$.

例 2a (模拟一个指数随机变量) 设 $F(x) = 1 - \mathrm{e}^{-x}$, 则 $F^{-1}(u)$ 是下列方程的解: $1 - \mathrm{e}^{-x} = u$ 或 $x = -\ln(1-u)$

因此, 若 U 为 $(0,1)$ 均匀随机变量, 则

$$F^{-1}(U) = -\ln(1-U)$$

的分布为指数分布, 期望为 1. 由于 $(1-U)$ 也是 $(0,1)$ 均匀随机变量, $-\ln U$ 也是指数随机变量, 其期望为 1. 若 X 具有指数分布, 其期望为 1, 则 cX 具有指数分布, 其期望为 c. 利用指数分布的这个特点知 $-c\ln U$ 具有指数分布, 期望为 c. ■

例 2a 的结果也可以用来模拟 Γ 随机变量.

例 2b (模拟一个 $\Gamma(n,\lambda)$ 随机变量)

当 n 为整数时, 我们可利用 Γ 随机变量与指数随机变量的关系模拟 $\Gamma(n,\lambda)$ 随机变量, $\Gamma(n,\lambda)$ 随机变量是 n 个相互独立同分布的指数随机变量的和, 每个和项具有期望 $1/\lambda$. 因此, 设 $U_1, \cdots, U_n$ 为随机数 (独立同分布的 $(0,1)$ 均匀随机变量), 则

$$X = -\sum_{i=1}^{n} \frac{1}{\lambda} \ln U_i = -\frac{1}{\lambda} \ln\left(\prod_{i=1}^{n} U_i\right)$$

具有 $\Gamma(n,\lambda)$ 分布. ■

10.2.2 舍取法

假定我们有办法模拟一个随机变量, 其密度函数为 g, 我们可以以这个随机变量为基础, 模拟一个密度为 f 的随机变量. 其方法是; 先模拟一个随机变量 Y, Y 的分布密度为 g, 然后以正比于 $f(Y)/g(Y)$ 的概率采用 Y 的值. 具体说来, 令 c 为一常数, 满足

$$\frac{f(y)}{g(y)} \leqslant c \qquad \text{对一切}y\text{成立}.$$

然后采用下列步骤产生具有密度 f 的随机变量.

第 1 步 模拟 Y, 使 Y 具有密度 g, 同时产生一随机数 U.

第 2 步 若 $U \leqslant f(Y)/[cg(Y)]$, 则 $X = Y$, 否则回到第一步.

舍取法模拟流程见图 10.1.

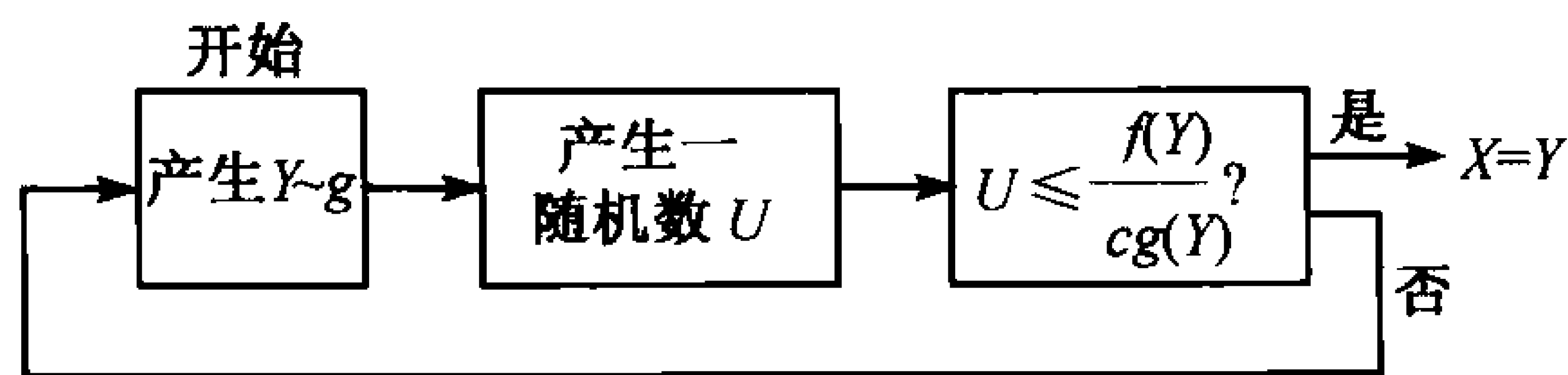


图 10.1 利用舍取法产生具有密度函数 f 的随机变量 X

下面命题说明舍取法的原理.

命题 2.2 由上述舍取法产生的随机变量具有密度函数 f.

证明: 设 X 为由舍取法产生的随机变量, 记 N 为舍取法中循环的次数, 则

$$P\{X \leqslant x\} = P\{Y_N \leqslant x\} = P\left\{Y \leqslant x | U \leqslant \frac{f(Y)}{cg(Y)}\right\}$$

$$= \frac{P\left\{Y \leqslant x, U \leqslant \dfrac{f(Y)}{cg(Y)}\right\}}{K}$$

其中 $K = P\{U \leqslant f(Y)/[cg(Y)]\}$. 由于 Y 与 U 相互独立, Y 与 U 的联合密度由下式给出:

$$f(y,u) = g(y) \qquad 0 < u < 1$$

这样,

$$P\{X \leqslant x\} = \frac{1}{K} \iint\limits_{\substack{y \leqslant x \\ 0 \leqslant u \leqslant f(y)/cg(y)}} g(y)\,\mathrm{d}u\,\mathrm{d}y$$

$$= \frac{1}{K}\int_{-\infty}^{x}\int_{0}^{f(y)/cg(y)} \mathrm{d}u\, g(y)\,\mathrm{d}y = \frac{1}{cK}\int_{-\infty}^{x} f(y)\,\mathrm{d}y \qquad (2.2)$$

由于 $f(y)$ 为密度函数, 上式两边令 $x \to +\infty$, 得

$$1 = \frac{1}{cK}\int_{-\infty}^{\infty} f(y)\mathrm{d}y = \frac{1}{cK}$$

这样, (2.2) 式变成

$$P\{X \leqslant x\} = \int_{-\infty}^{x} f(y)\,\mathrm{d}y$$

□

注释 (a) 前面提到以概率 $f(Y)/[cg(Y)]$ 接受 Y 是指产生一个随机数 U, 若 $U \leqslant f(Y)/[cg(Y)]$, 则令 $X = Y$.

(b) 在产生随机数的过程中, 接收 Y 的概率为 $P\{U \leqslant f(Y)/[cg(Y)]\} = K = 1/c$. 由此可知, 循环次数 N 的分布是以 c 为期望的几何分布. ■

例 2c (模拟一个正态随机变量)

设 Z 是一个标准正态随机变量 (期望为 0, 方差为 1). 注意 $X = |Z|$ 具有密度

函数

$$f(x) = \frac{2}{\sqrt{2\pi}} e^{-x^2/2} \qquad 0 < x < \infty \tag{2.3}$$

开始时, 我们先用舍取法, 模拟 X. 取一个密度函数 $g(x)$

$$g(x) = e^{-x} \qquad 0 < x < \infty$$

注意到

$$\begin{aligned}\frac{f(x)}{g(x)} &= \sqrt{\frac{2}{\pi}} \exp\left\{\frac{-(x^2 - 2x)}{2}\right\} = \sqrt{\frac{2}{\pi}} \exp\left\{\frac{-(x^2 - 2x + 1)}{2} + \frac{1}{2}\right\} \\ &= \sqrt{\frac{2e}{\pi}} \exp\left\{\frac{-(x-1)^2}{2}\right\} \leqslant \sqrt{\frac{2e}{\pi}}\end{aligned} \tag{2.4}$$

取 $c = \sqrt{2e/\pi}$, 由 (2.4) 式知

$$\frac{f(x)}{cg(x)} = \exp\left\{\frac{-(x-1)^2}{2}\right\}$$

现在, 我们可以用舍取法模拟 X:

(a) 产生独立随机变量 Y 和 U, 其中 Y 具有期望为 1 的指数分布, U 为 $(0,1)$ 上均匀分布.

(b) 若 $U \leqslant \exp\{-(Y-1)^2/2\}$, 则 $X = Y$, 否则转向 (a).

当 X 得到以后 [X 具有密度 (2.3)], 可令 $Z = +X$ 或 $-X$, 以 1/2 的概率取正号, 1/2 的概率取负号.

在步骤 (b) 中, 条件 $U \leqslant \exp\{-(Y-1)^2/2\}$ 等价于 $-\ln U \geqslant (Y-1)^2/2$, 但是在例 2a 中指出, $-\ln U$ 是指数随机变量, 期望为 1, 因此, 步骤 (a),(b) 等价于:

(a′) 产生两个相互独立的, 期望为 1 的指数随机变量 Y_1 和 Y_2.

(b′) 若 $Y_2 \geqslant (Y_1 - 1)^2/2$, 令 $X = Y_1$, 否则转向 (a′).

当 Y_1 被接受时, Y_2 必大于 $(Y_1 - 1)^2/2$, 但是 Y_2 比 $(Y_1 - 1)^2/2$ 大多少? 利用指数分布的无记忆性, 当 Y_2 值超过某值时, 超过部分的时间是一个均值为 1 的指数随机变量, 因此, $Y_2 - (Y_1 - 1)^2/2$ 也是指数分布随机变量, 其均值为 1. 这样, 在得到 X 的同时, 我们也能得到另一个指数随机变量 $Y_2 - (Y_1 - 1)^2/2$, 它与 X 相互独立, 并且期望为 1. 这样, 在模拟标准正态随机变量时, 我们还附加地产生一个指数随机变量.

综合起来, 我们可以利用下列步骤产生一个指数随机变量 (期望为 1) 和与之独立的标准正态随机变量.

第 1 步 产生 Y_1, 指数随机变量, 期望为 1.

第 2 步 产生 Y_2, 指数随机变量, 期望为 1.

第 3 步 若 $Y_2 - (Y_1 - 1)^2/2 > 0$, 令 $Y = Y_2 - (Y_1 - 1)^2/2$, 转向第 4 步, 否则转向第一步.

第 4 步 产生一个随机数 U, 令

$$Z = \begin{cases} Y_1 & U \leqslant \frac{1}{2} \\ -Y_1 & U > \frac{1}{2} \end{cases}$$

上述过程所得到的 Y 与 Z 相互独立, Z 为标准正态随机变量, Y 为指数随机变量, 期望为 1(若要求得正态随机变量, 其期望为 μ, 方差为 σ^2 只需取 $\mu + \sigma Z$.).

注释 (a) 由于 $c = \sqrt{2e/\pi} \approx 1.32$, 在前述的产生随机变量过程中, 第 3 步中要求有 N 步的循环, 其中 N 是均值 $c \approx 1.32$ 的几何随机变量.

(b) 若我们希望产生一个标准正态的随机变量序列, 则在第 3 步中生产的 Y 可供产生下一个正态随机变量之用. 这样算来, 产生一个正态随机变量需要 1.64($= 2 \times 1.32 - 1$) 个指数随机变量以及 1.32 次平方运算. ■

例 2d (模拟正态随机变量 —— 极坐标法) 在第 6 章例 7b 中指出, 若 X, Y 为相互独立的标准正态随机变量, 则它们的极坐标 $R = \sqrt{X^2 + Y^2}$ 和 $\Theta = \arctan(Y/X)$ 相互独立, R^2 为均值 2 的指数随机变量. Θ 在 $(0, 2\pi)$ 上均匀分布, 因此, 若取 U_1, U_2为随机数 (利用例 2a 的方法生成), 我们得到

$$R = (-2\ln U_1)^{1/2} \qquad \Theta = 2\pi U_2$$

从而

$$X = R\cos\Theta = (-2\ln U_1)^{\frac{1}{2}}\cos(2\pi U_2) \quad Y = R\sin\Theta = (-2\ln U_1)^{\frac{1}{2}}\sin(2\pi U_2) \quad (2.5)$$

为独立的标准正态随机变量. ■

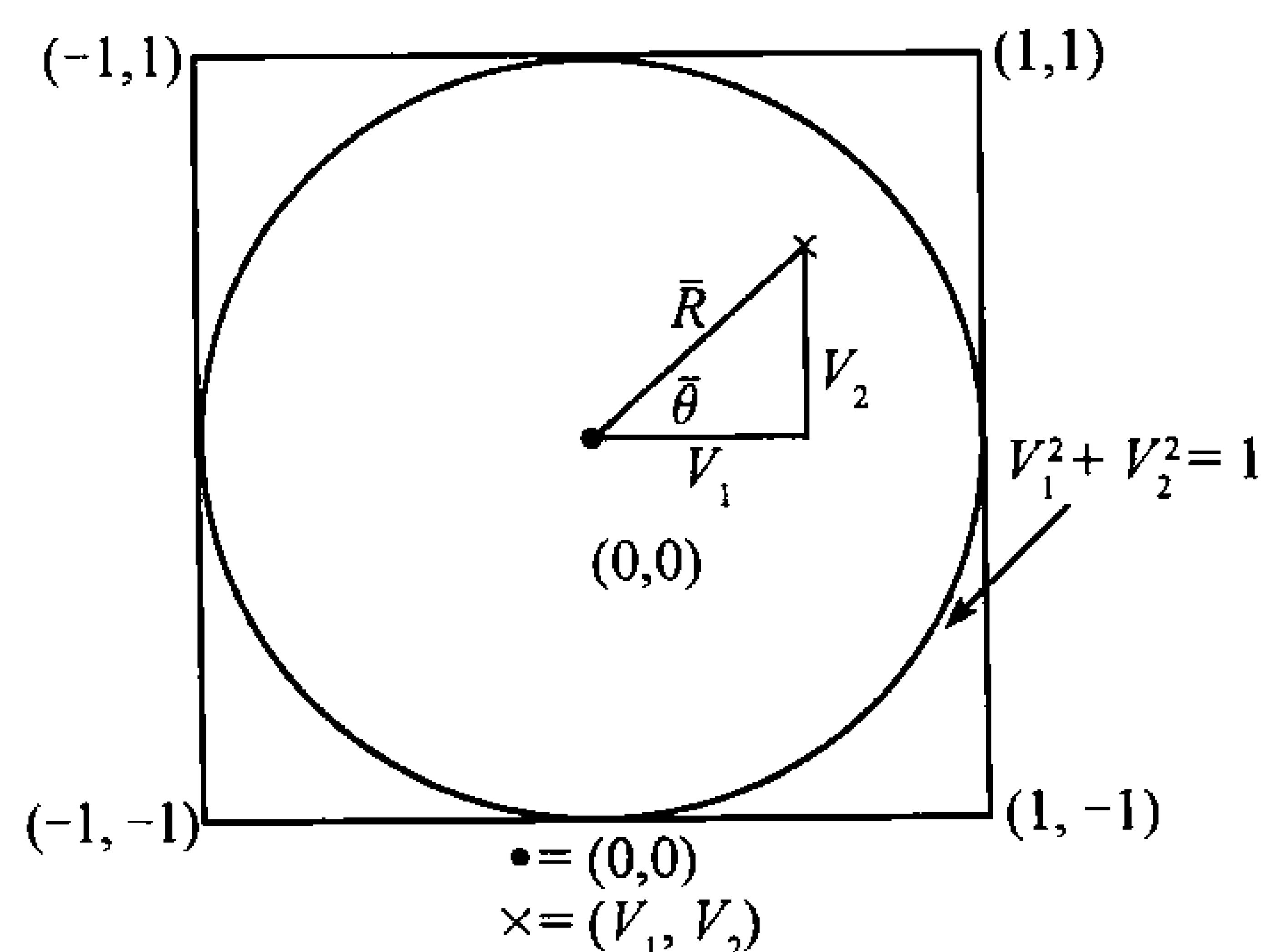


图 10.2 极坐标法模拟正态随机变量

上面提供的产生标准正态随机变量的方法称为 Box-Muller 方法, 由于需要计算 $\sin\Theta$ 与 $\cos\Theta$, 其效率受到质疑, 但是可用下面方法避免费时的困难, 由于 U 为 $(0,1)$ 均匀随机变量, $2U$ 为 $(0,2)$ 均匀随机变量, 所以 $2U - 1$ 是 $(-1, +1)$ 上的均匀随机变量. 设 U_1, U_2 为两个随机数, 令

$$V_1 = 2U_1 - 1 \qquad V_2 = 2U_2 - 1$$

则 (V_1, V_2) 是在面积为 4, 中心为 $(0,0)$ 的一个方块上均匀分布的随机向量.

我们假定连续产生 (V_1, V_2), 直到产生一对 (V_1, V_2) 满足条件 $V_1^2 + V_2^2 \leqslant 1$, 即 (V_1, V_2) 处于以 $(0,0)$ 为中心的单位圆内 (见图 10.2), 显然这样得到的 (V_1, V_2) 是在单位圆内均匀分布的随机向量. 取 $(\bar{R}, \bar{\Theta})$ 为 V_1, V_2 的极坐标. 容易验证 $\bar{R}$ 与 $\bar{\Theta}$ 相

互独立, $\bar{R}^2$ 为 $(0,1)$ 均匀随机变量, $\bar{\Theta}$ 为 $(0,2\pi)$ 上均匀随机变量 (见习题13). 由于

$$\sin\bar{\Theta}=\frac{V_2}{\bar{R}}=\frac{V_2}{\sqrt{V_1^2+V_2^2}}\qquad \cos\bar{\Theta}=\frac{V_1}{\bar{R}}=\frac{V_1}{\sqrt{V_1^2+V_2^2}}$$

由 (2.5) 式知, 标准正态随机变量 X,Y 可从下式得到,

$$X=(-2\ln U)^{1/2}V_1/\bar{R}\qquad Y=(-2\ln U)^{1/2}V_2/\bar{R}$$

事实上, 由于 (在 $V_1^2+V_2^2\leqslant 1$ 的条件下)$\bar{R}^2$ 为 $(0,1)$ 均匀随机变量, 与 $\bar{\Theta}$ 独立, 这样, $\bar{R}^2$ 可代替 U 而不必重新产生新的随机数.

$$X=(-2\ln\bar{R}^2)^{1/2}\frac{V_1}{\bar{R}}=\sqrt{\frac{-2\ln S}{S}}V_1\qquad Y=(-2\ln\bar{R}^2)^{1/2}\frac{V_2}{\bar{R}}=\sqrt{\frac{-2\ln S}{S}}V_2$$

为相互独立的标准正态随机变量, 其中

$$S=\bar{R}^2=V_1^2+V_2^2$$

综合起来, 利用下列方法可以产生一对独立的标准正态随机变量.

第 1 步 产生随机数 U_1,U_2;

第 2 步 令 $V_1=2U_1-1,V_2=2U_2-1,S=V_1^2+V_2^2$;

第 3 步 若 $S>1$, 转向第 1 步;

第 4 步 得到独立的标准正态随机变量

$$X=\sqrt{\frac{-2\ln S}{S}}V_1\qquad Y=\sqrt{\frac{-2\ln S}{S}}V_2$$

上面的方法称为极坐标法, 由于在正方形中随机点落入单位圆内的概率为 $\pi/4$ (圆面积与正方形面积之比), 平均来说要经过 $4/\pi\approx 1.273$ 次循环产生 2 个相互独立的标准正态随机变量. 因此, 平均来说要求 2.546 个随机数, 一次求对数, 一次求平方根, 一次除法, 4.546 次乘法才能求得两个相互独立的正态随机数.

例 2e(模拟一个 χ^2 随机变量) n 个自由度的 χ^2 分布就是随机变量 $\chi_n^2=Z_1^2+\cdots+Z_n^2$ 的分布, 其中 $Z_1,\cdots,Z_n$ 是独立标准正态分布的随机变量序列, 在第 6 章6.3 节指出, $Z_1^2+Z_2^2$ 具有指数分布, 期望为 2, 因此当 n 为偶数时, 例如 $n=2k$, 则 χ_{2k}^2 的分布为 Γ 分布, 参数为 $(k,1/2)$.

这样, $-2\ln(\prod_{i=1}^k U_i)$ 具有 $2k$ 个自由度的 χ^2 分布, 当 $n=2k+1$ 为奇数时, 我们只需先求出一个标准正态随机变量 Z, 再将 Z^2 加到 χ_{2k}^2 可得到 $2k+1$ 个自由度的 χ^2 随机变量, 即

$$\chi_{2k+1}^2=Z^2-2\ln\left(\prod_{i=1}^k U_i\right)$$

其中 $Z,U_1,\cdots,U_k$ 相互独立, $U_1,\cdots,U_k$ 为随机数序列, Z 为标准正态随机变量. ■

10.3 模拟离散分布

模拟具有连续分布函数的随机变量所使用的方法, 都可适用于离散随机变量的模拟. 例如, 我们希望模拟具有下列分布列的随机变量 X:

$$P\{X=x_j\}=P_j \qquad j=0,1,\cdots,\sum_j P_j=1$$

可利用下面的方法, 它是反变换方法的离散版本.

设 U 为随机数, 令

$$X=\begin{cases} x_1 & 若\ U\leqslant P_1 \\ x_2 & 若\ P_1<U\leqslant P_1+P_2 \\ \cdots & \\ x_j & 若\ \sum\limits_{i=1}^{j-1}P_i<U\leqslant\sum\limits_{i=1}^{j}P_i \\ \cdots & \end{cases}$$

由于

$$P\Big\{X=x_j\Big\}=P\left\{\sum_{i=1}^{j-1}P_i<U\leqslant\sum_{i=1}^{j}P_i\right\}=P_j$$

我们可以看出, 所产生的随机变量 X 具有离散分布列 $\{P_j,j=1,2,\cdots\}$.

例 3a (几何分布) 设有一独立重复试验, 每次成功的概率为 $p, 0<p<1$, 试验一直到出现成功为止, 记 X 为试验的次数, 则

$$P\{X=i\}=(1-p)^{i-1}p \qquad i\geqslant 1$$

$X=i$ 表示前 $i-1$ 次试验的均为失败, 而第 i 次成功. 随机变量 X 的分布是参数为 p 的几何分布. 由于

$$\begin{aligned}&\sum_{i=1}^{j-1}P\{X=i\}=1-P\{X>j-1\}\\ &=1-P\{前j-1次试验均为失败\}=1-(1-p)^{j-1} \qquad j\geqslant 1\end{aligned}$$

这样, X 可以由下列方式产生: 取 U 为随机数 [即 $(0,1)$ 上均匀随机变量, 下同], 当

$$1-(1-p)^{j-1}<U\leqslant 1-(1-p)^j$$

时, X 取为 j, 上式与

$$(1-p)^j\leqslant 1-U<(1-p)^{j-1}$$

是等价的, 又由于 U 与 $1-U$ 具有相同的分布, 因此, X 具有下面的表达式

$$X = \min\{j : (1-p)^j \leqslant U\} = \min\{j : j\ln(1-p) \leqslant \ln U\} = \min\left\{j : j \geqslant \frac{\ln U}{\ln(1-p)}\right\}$$

上式中有一个不等号反向的过程, 其原因是, $\ln(1-p) < \ln 1 = 0$, 用这个数去除不等号的两边时, 不等号应反向. 利用记号 $[x]$($[x]$ 为不超过 x 的最大整数), X 可以写成

$$X = 1 + \left[\frac{\ln U}{\ln(1-p)}\right]$$ ■

和连续情形类似, 对于某些离散分布, 也开发出某些模拟方法. 现举出其中 2 个方法.

例 3b (模拟一个二项随机变量) 二项随机变量［参数为 (n,p)］可以表示 n 个独立的伯努利随机变量之和, 利用这一点, 很容易进行模拟, 设 $U_1, \cdots, U_n$ 为一组随机数, 令

$$X_i = \begin{cases} 1 & 若\ U_i < p \\ 0 & 其他 \end{cases}$$

易知 $X \equiv \sum_{i=1}^n X_i$ 是二项随机变量, 其参数为 (n,p). ■

例 3c (模拟一个泊松随机变量) 设 $U_1, U_2, \cdots$ 为一串随机数, 记 $N = \min\{n : \prod_{i=1}^n U_i < \mathrm{e}^{-\lambda}\}$, 我们将指出随机变量 $X \equiv N-1$ 具有泊松分布, 其期望为 λ. 注意到

$$X + 1 = \min\left\{n : \prod_{i=1}^n U_i < \mathrm{e}^{-\lambda}\right\}$$

与

$$X = \max\left\{n : \prod_{i=1}^n U_i \geqslant \mathrm{e}^{-\lambda}\right\} \quad 其中 \prod_{i=1}^0 U_i \equiv 1$$

是等价的. 上式经过化简得

$$X = \max\left\{n : \sum_{i=1}^n \ln U_i \geqslant -\lambda\right\} = \max\left\{n : \sum_{i=1}^n -\ln U_i \leqslant \lambda\right\}$$

注意到 $-\ln U_i$ 具有指数分布, 其期望值为 1. 现在考虑一个泊松过程, 其速率为 1, 易知, 这个过程在 $(0,\lambda)$ 上事件的个数的分布为泊松分布, 其期望为 λ, 而 $(0,\lambda)$ 上每两个相邻事件之间的时间间隔刚好为参数为 1 的指数分布, 而且这些时间间隔又相互独立. 现在再看一看 X 的表达式, 其分布刚好与泊松过程在 $(0,\lambda]$ 上事件的个数的分布相同. 因此, X 的分布为泊松分布, 其期望为 λ. ■

10.4 方差缩减技术

设 $X_1, \cdots, X_n$ 具有一已知的联合分布, 现在我们希望计算

$$\theta \equiv E[g(X_1, \cdots, X_n)]$$

其中 g 是一个已知的函数, 有时候要解析地计算这个值是十分困难的. 但我们可以利用模拟技术去估计 θ 的值. 我们可以产生一组 $X_1^{(1)}, \cdots, X_n^{(1)}$, 使得 $X_1^{(1)}, \cdots, X_n^{(1)}$ 与 $X_1, \cdots, X_n$ 具有相同的联合分布, 令

$$Y_1 = g(X_1^{(1)}, \cdots, X_n^{(1)})$$

同时, 我们还可以模拟一组 $X_1^{(2)}, \cdots, X_n^{(2)}$, 使得它与第一组变量相互独立并且具有相同的联合分布, 令

$$Y_2 = g(X_1^{(2)}, \cdots, X_n^{(2)})$$

继续模拟下去, 得到 $Y_1, Y_2, \cdots, Y_k$(k 为事先确定), $Y_1, \cdots, Y_k$ 相互独立, 且与 $g(X_1, \cdots, X_n)$ 分布相同, 记 $\bar{Y}$ 为 $Y_1, \cdots, Y_k$ 的平均数, 即

$$\bar{Y} = \frac{1}{k}\sum_{i=1}^{k} Y_i$$

则

$$E[\bar{Y}] = \theta \qquad E[(\bar{Y} - \theta)^2] = \mathrm{Var}(\bar{Y})$$

由上式可知, $\bar{Y}$ 的均方误差等于 $\bar{Y}$ 的方差, 我们希望 $\mathrm{Var}(\bar{Y})$ 越小越好. [在我们所讨论的情况下, $\mathrm{Var}(\bar{Y}) = \frac{1}{k}\mathrm{Var}(Y_1)$, 而通常这个量是不知道的, 我们必须设法从 $Y_1, \cdots, Y_k$ 的值去估计它.] 现在我们需要介绍几个方差缩减技术.

10.4.1 利用对偶变量

前面我们介绍估计 θ 的方法时, 假定所产生的 Y_1 与 Y_2 相互独立且同分布. 现在讨论关于方差的一个公式

$$\begin{aligned}\mathrm{Var}\left(\frac{Y_1 + Y_2}{2}\right) &= \frac{1}{4}[\mathrm{Var}(Y_1) + \mathrm{Var}(Y_2) + 2\mathrm{Cov}(Y_1, Y_2)] \\ &= \frac{1}{2}\mathrm{Var}(Y_1) + \frac{1}{2}\mathrm{Cov}(Y_1, Y_2)\end{aligned}$$

由上式可以看出, 若 Y_1 与 Y_2 具有负相关, 则从缩减方差的角度看, 可将 $1/2(Y_1 + Y_2)$ 的方差压缩. 现在看一看如何实现把方差缩减. 先假定 $X_1, \cdots, X_n$ 是相互独立的, 此外, 每一个变量都是利用反变换方法产生的, 即 $X_i = F_i^{-1}(U_i)$, 其中 U_i 是随机数, F_i 是 X_i 的分布函数, Y_1 可表示成

$$Y_1 = g(F_1^{-1}(U_1), \cdots, F_n^{-1}(U_n))$$

由于 $1 - U$ 也是 $(0, 1)$ 随机数, $1 - U$ 与 U 具有负相关, 若定义

$$Y_2 = g(F_1^{-1}(1 - U_1), \cdots, F_n^{-1}(1 - U_n))$$

则 Y_2 与 Y_1 同分布. 因此, 若 Y_1, Y_2 负相关, 那么 $(Y_1 + Y_2)/2$ 的方差就会比 $\frac{1}{2}\mathrm{Var}(Y_1)$

小. (另外, 从计算的角度, 也节省了计算量, 在产生 Y_2 的时候, 本来应该产生 n 个新的随机数, 现在可用 $1-U_i, i=1,\cdots,n$ 代之.) 在一般情况下, 我们不能证明 Y_1 和 Y_2 为负相关, 但通常情况下, 它们是负相关的, 特别当 g 是单调函数时, 可以证明它们是负相关的.

10.4.2 利用"条件"缩减方差

首先回忆下列条件方差公式 (见 7.5.4 节):

$$\mathrm{Var}(Y)=E[\mathrm{Var}(Y|Z)]+\mathrm{Var}(E[Y|Z])$$

现在假定我们要估计 $E[g(X_1,\cdots,X_n)]$, 先模拟 $\boldsymbol{X}=(X_1,\cdots,X_n)$, 再计算 $Y=g(\boldsymbol{X})$, 但是, 若存在某随机变量 Z, 并且能够计算 $E[Y|Z]$, 由于 $\mathrm{Var}(Y|Z)\geqslant 0$, 利用条件方差公式得

$$\mathrm{Var}(E[Y|Z])\leqslant \mathrm{Var}(Y)$$

由于 $E[E[Y|Z]]=E[Y]$, 可知 $E[Y|Z]$ 比 Y 更好.

例 4a (估计 π) 令 U_1,U_2 为随机数, 记 $V_i=2U_i-1, i=1,2$, 在 例 2d 中已经指出 (V_1,V_2) 在面积为 4, 中心为 $(0,0)$ 的一个方块内均匀分布, 随机点 (V_1,V_2) 落在半径为 1 的以 $(0,0)$ 为圆心的圆内的概率为 $\pi/4$ (见图 10.2, $\pi/4$ 等于内接圆与正方形面积之比). 现在可模拟 (V_1,V_2) n 次, 令

$$I_j=\begin{cases}1 & \text{第 } j \text{ 次模拟落在单位圆内}\\ 0 & \text{其他}\end{cases}$$

故 $I_1,\cdots,I_n$ 为独立同分布, $E[I_j]=\pi/4$, 利用强大数定律,

$$\frac{I_1+\cdots+I_n}{n}\to\frac{\pi}{4}\qquad n\to\infty$$

将这个比例乘以 4, 可得 π 的估计.

利用条件期望可改善这个估计, 对于示性函数 I, 考虑条件概率

$$E[I|V_1]=P\{V_1^2+V_2^2\leqslant 1|V_1\}=P\{V_2^2\leqslant 1-V_1^2|V_1\}$$

由于

$$\begin{aligned}&P\{V_2^2\leqslant 1-V_1^2|V_1=v\}=P\{V_2^2\leqslant 1-v^2\}\\ =&P\{-\sqrt{1-v^2}\leqslant V_2\leqslant\sqrt{1-v^2}\}=\sqrt{1-v^2}\end{aligned}$$

即 $E[I|V_1]=\sqrt{1-V_1^2}$, 因此, 利用 $\sqrt{1-V_1^2}$ 的平均值作为 $\pi/4$ 的估计是原来的估计的改进. 又由于

$$E\left[\sqrt{1-V_1^2}\right]=\int_{-1}^{1}\frac{1}{2}\sqrt{1-v^2}\mathrm{d}v=\int_0^1\sqrt{1-u^2}\mathrm{d}u=E\left[\sqrt{1-U^2}\right]$$

其中 U 为随机数, 即 $(0,1)$ 上均匀分布的随机变量, 因此, 我们模拟产生 n 个随机数, 利用 $\sqrt{1-U^2}$ 的平均值作为 $\pi/4$ 的估计, 具有更高的精度 (习题 14 指出, 利用 $\sqrt{1-V^2}$ 产生的估计与利用 $\sqrt{1-v^2}$ 产生的估计具有相同的精度).

由于 $g(u)=\sqrt{1-u^2}, 0\leqslant u\leqslant 1$ 是一个单调递减函数, 我们可以利用对偶变量法, 进一步减少估计的误差, 我们可用 $n/2$ 个 $(\sqrt{1-U^2}+\sqrt{1-(1-U)^2})/2$ 的平均值来估计 $\pi/4$, 下面的表列出了 $n=10\ 000$ 时, π 的估计值.

方　　法	π的估计
落入单位圆内随机点的比例	3.1612
$\sqrt{1-U^2}$ 的平均值	3.128448
$\frac{1}{2}(\sqrt{1-U^2}+\sqrt{1-(1-U^2)})$ 的平均值	3.139578

利用最后一个方法, 当 $n=64\ 000$ 时, π 的估值为 3.143 288. ■

10.4.3 控制变量

假设我们希望模拟 $E[g(\boldsymbol{X})]$, 其中 $\boldsymbol{X}=(X_1,\cdots,X_n)$. 但是我们已知某 $f(\boldsymbol{X})$ 的期望, 例如 $E[f(\boldsymbol{X})]=\mu$, 我们可利用

$$W=g(\boldsymbol{X})+a[f(\boldsymbol{X})-\mu]$$

来模拟 $E[g(\boldsymbol{X})]$, 但

$$\mathrm{Var}(W)=\mathrm{Var}[g(\boldsymbol{X})]+a^2\mathrm{Var}[f(\boldsymbol{X})]+2a\mathrm{Cov}[g(\boldsymbol{X}),f(\boldsymbol{X})] \tag{4.1}$$

在上式中令

$$a=\frac{-\mathrm{Cov}[f(\boldsymbol{X}),g(\boldsymbol{X})]}{\mathrm{Var}[f(\boldsymbol{X})]} \tag{4.2}$$

得

$$\mathrm{Var}(W)=\mathrm{Var}[g(\boldsymbol{X})]-\frac{\mathrm{Cov}[f(\boldsymbol{X}),g(\boldsymbol{X})]^2}{\mathrm{Var}[f(\boldsymbol{X})]} \tag{4.3}$$

但是通常, $\mathrm{Var}[f(\boldsymbol{X})]$ 和 $\mathrm{Cov}[f(\boldsymbol{X}),g(\boldsymbol{X})]$ 是未知的, 因此, 我们得不到所需的方差缩减. 实践中, 我们可以利用模拟数据去估计这个值. 理论上我们可以利用这个方法对所有的模拟结果缩减相应的方差.

小　　结

设 F 是一个连续的分布函数, U 是 $(0,1)$ 上的均匀分布的随机变量 (称为随机数) 则 $F^{-1}(U)$ 具有分布 F, 其中 $F^{-1}(u)$ 是方程 $F(x)=u$ 的解, 这种由随机数构造其他随机变量的方法称为反变换方法.

另一个产生随机变量的方法称为舍取法. 假定对于密度 g, 我们已经有一个产生随机变量的成熟的方法, 现在希望模拟一个具有密度 f 的随机变量. 我们首先确

定一个常数 c, 它满足

$$\max \frac{f(x)}{g(x)} \leqslant c$$

然后经过下列步骤:

第 1 步 产生 Y, 其密度为 g.

第 2 步 产生随机数 U.

第 3 步 若 $U \leqslant f(Y)/[cg(Y)]$, 则令 $X=Y$, 过程中止.

第 4 步 回到第 1 步. 此方法循环的次数具有几何分布, 其平均值为 c.

标准正态随机变量可通过舍取法产生 (g 为指数密度, 期望为 1) 或者利用极坐标方法产生.

为了估计某一个参数 θ, 首先模拟一个随机变量, 使得它的期望值为 θ, 然后, 利用统计方法缩减其相应的方差. 本文中介绍了三种缩减方差的方法:

1. 利用对偶变量.

2. 利用条件期望.

3. 利用控制变量.

习　题

1. 下面的方法可能产生 $\{1,2,\cdots,n\}$ 的一个随机排列. 这个方法比例 1a 中介绍的方法快, 但是这个方法一直到计算结束, 才把每个位置上的元素确定下来. 此处 $P(i)$ 表示位置 i 上的元素.

   第 1 步 令 $k=1$.
   第 2 步 令 $P(1)=1$.
   第 3 步 若 $k=n$, 则停止, 否则令 $k=k+1$.
   第 4 步 产生一随机数 U, 令

   $$P(k)=P([kU]+1) \qquad P([kU]+1)=k$$

   回到第 3 步.

   (a) 解释为什么这个方法行得通.

   (b) 指出在 k 步, 即 $P(k)$ 确定以后, $P(1),\cdots,P(k)$ 是 $1,2,\cdots,k$ 的一个随机排列.

   提示: 利用归纳法, 并且指出

   $$P_k\{i_1,i_2,\cdots,i_{j-1},k,i_j,\cdots,i_{k-2},i\}=P_{k-1}\{i_1,i_2,\cdots,i_{j-1},i,i_j,\cdots,i_{k-2}\}\frac{1}{k}$$
   $$=\frac{1}{k!} \qquad \text{利用归纳法}$$

2. 模拟一个随机变量具有下列密度:

   $$f(x)=\begin{cases}\mathrm{e}^{2x} & -\infty<x<0\\ \mathrm{e}^{-2x} & 0<x<\infty\end{cases}$$

3. 模拟一个随机变量具有下列密度:

$$f(x)=\begin{cases}\frac{1}{2}(x-2) & 2\leqslant x\leqslant 3\\ \frac{1}{2}\left(2-\frac{x}{3}\right) & 3<x\leqslant 6\\ 0 & \text{其他}\end{cases}$$

4. 找出一个模拟具有下列分布函数的随机变量的方法:

$$F(x)=\begin{cases}0 & x\leqslant -3\\ \frac{1}{2}+\frac{x}{6} & -3<x<0\\ \frac{1}{2}+\frac{x^2}{32} & 0<x\leqslant 4\\ 1 & x>4\end{cases}$$

5. 利用反变换产生一个威布尔分布的随机变量, 威布尔分布的分布函数由下式给出:

$$F(t)=1-\mathrm{e}^{-at^{\beta}} \qquad t\geqslant 0$$

6. 给出一个模拟具有下面失效率函数的随机变量:

(a) $\lambda(t)=c$; (b) $\lambda(t)=ct$; (c) $\lambda(t)=ct^2$; (d) $\lambda(t)=ct^3$.

7. 设 F 是一个分布函数

$$F(x)=x^n \qquad 0<x<1$$

(a) 只利用一个随机数, 给出一个模拟此分布的方法.

(b) 记 $U_1,\cdots,U_n$ 为独立随机数, 指出

$$P\{\max(U_1,\cdots,U_n)\leqslant x\}=x^n$$

(c) 利用 (b), 给出一个模拟 F 的方法.

8. 假定模拟 $F_i, i=1,\cdots,n$ 是比较容易的, 如何模拟下列分布函数?

(a) $F(x)=\prod_{i=1}^{n}F_i(x)$. (b) $F(x)=1-\prod_{i=1}^{n}[1-F_i(x)]$.

9. 假定我们有一个方法模拟 F_1 和 F_2, 如何模拟

$$F(x)=pF_1(x)+(1-p)F_2(x) \qquad 0<p<1$$

给出一个模拟下列分布的方法:

$$F(x)=\begin{cases}\frac{1}{3}(1-\mathrm{e}^{-3x})+\frac{2}{3}x & 0<x\leqslant 1\\ \frac{1}{3}(1-\mathrm{e}^{-3x})+\frac{2}{3} & x>1\end{cases}$$

10. 在例 2c 中, 利用舍取法模拟标准正态随机变量时利用了指数分布 ($\lambda=1$, 期望为 $\frac{1}{\lambda}=1$). 现在的问题是: 是否可以利用其他的指数密度, 达到更高的效率, 例如利用密度 $g(x)=\lambda\mathrm{e}^{-\lambda x}$, $x>0$. 指出平均循环次数当 $\lambda=1$ 时达到最小.

11. 利用舍取法, $g(x)=1, 0<x<1$, 求一个模拟下列分布的算法:

$$f(x)=\begin{cases}60x^3(1-x)^2 & 0<x<1\\ 0 & 其他\end{cases}$$

12. 怎样利用随机数去逼近 $\int_0^1 k(x)\mathrm{d}x$, 其中 $k(x)$ 是任一函数?

提示: 若 U 是随机数, $E[k(U)]$ 是什么?

13. 设 (X,Y) 为以 $(0,0)$ 为圆心, 半径为 1 的圆上均匀分布, 它的密度为

$$f(x,y)=\frac{1}{\pi} \qquad 0\leqslant x^2+y^2\leqslant 1$$

记 $R=(x^2+y^2)^{1/2}$ 和 $\theta=\arctan(Y/X)$ 表示它的极坐标. 指出 R 与 θ 相互独立, R^2 为 $(0,1)$ 上均匀随机变量, θ 为 $(0,2\pi)$ 上均匀随机变量.

14. 在例 4a 中, 我们已指出

$$E[(1-V^2)^{1/2}]=E[(1-U^2)^{1/2}]=\frac{\pi}{4}$$

其中 V 在 $(-1,1)$ 上均匀分布, 而 U 在 $(0,1)$ 上均匀分布, 指出

$$\mathrm{Var}[(1-V^2)^{1/2}]=\mathrm{Var}[(1-U^2)^{1/2}]$$

15. (a) 验证: 当 a 由 (4.2) 给出时, (4.1) 达到极小值.

(b) 验证 (4.1) 的极小值由 (4.3) 给出.

16. 设 X 取值于 $(0,1)$, 其密度为 $f(x)$. 指出模拟 $g(X)/f(X)$ 可以估计 $\int_0^1 g(x)\mathrm{d}x$. 这个方法称为重要样本法, 其要点是选择 f 与 g 相似, 使得 $g(X)/f(X)$ 具有较小的方差.

自 检 习 题

1. 设 X 具有概率密度

$$f(x)=C\mathrm{e}^x \qquad 0<x<1$$

(a) 找出常数 C; (b) 指出模拟 X 的方法.

2. 找出一个模拟随机变量的方法, 该随机变量具有密度

$$f(x)=30(x^2-2x^3+x^4) \qquad 0<x<1$$

3. 找出一个模拟离散随机变量的有效算法, 其分布列为

$$p_1=0.15 \qquad p_2=0.2 \qquad p_3=0.35 \qquad p_4=0.30$$

4. 设 X 是一个正态随机变量, 其期望为 μ, 方差为 σ^2. 定义一随机变量 Y, 使它与 X 具有相同的分布, 但是负相关.

5. 设 X,Y 具有独立指数随机变量, 其期望为 1.

(a) 利用模拟方法找出估计 $E[\mathrm{e}^{XY}]$ 的方法.

(b) 利用一个控制变量将 (a) 中得到的估计改进.

参 考 文 献

Ross, S. M. *Simulation*. 4th ed. San Diego, Calif.: Academic Press, Inc., 2006.[①]

① 本书中文版和英文影印版均已由人民邮电出版社出版. —— 编者注

索 引

Y

Z

其 他